The Journal
of the
Acoustical Society of America

(Continued)

CONTENTS—*Continued from preceding page*

CONTENTS—*Continued from preceding page*

CONTENTS—*Continued from preceding page*

FORUM

The First Joint Meeting of the Acoustical Societies of America and Japan—How it happened

John C. Burgess[a]
Department of Mechanical Engineering, University of Hawaii, 2540 Dole Street, Honolulu, Hawaii 96822

To the editor:

The First Joint Meeting of the Acoustical Societies of America and Japan was a major milestone in the history of both societies. It opened the doors for much productive research collaboration and many personal associations. Neither Society has been the same since. This is the story of how it happened.

It was a cold 1973 winter day in Ottawa, at the National Research Council of Canada. I was an Invited Researcher in the Acoustics Section of the Physics Department, on Sabbatical leave from the University of Hawaii. During one of the Section's coffee hours, Edgar Shaw, out of the blue, asked me, "When are we going to have a meeting in Hawaii?" It was an amusing thought, since Ottawa was then a frozen wonderland. It would have been easy to forget—except that it came from the President of the Acoustical Society of America.

Edgar's question gave birth to what became the First Joint Meeting of the Acoustical Society of America (ASA) and the Acoustical Society of Japan (ASJ), held at the Sheraton-Waikiki in Honolulu, Hawaii, 27 November–1 December 1978, with James L. Flanagan as ASA President and T. Nimura as ASJ President.

Edgar's thought germinated until my return to Hawaii. With "Why not?" in mind, I sent an offer to Betty Goodfriend. Betty's response was "the Executive Council will never approve it." To her surprise, the Council did, and appointed me the Meeting Chairman (MC).

At that time, a MC was responsible for everything that happened—or could happen—or should have happened—at the meeting location. There was no Executive Director or Meetings Committee. The MC reported directly to the Executive Council. It was a recognized but unwritten rule that the MC had authority to make all final decisions, even to over-rule a Society officer. In effect, the MC was the meeting CEO. Of course, the MC had to face the music if the decisions ruffled feathers.

A worrying thought gradually surfaced. Would company managers in the United States regard a Hawaii meeting as a boondoggle? The potential effect of this bias was all too clear. How could we marginalize it? If the meeting were a joint meeting with the ASJ, managers would be faced with missing out on what their Japanese competitors were doing.

It worked! Later, a Bell Labs colleague told me that one of the Lab managers had gone around saying, "They have *no right* to hold a meeting in Hawaii!" But Bell Labs acousticians were there.

A joint meeting required the agreement of the ASJ. But we were handicapped. As a Society, we knew almost nothing about the ASJ. The ASA office had only an address in Tokyo. ASA President Bob Gales wrote to the ASJ president with the suggestion. The ASJ agreed, and appointed Professor Juichi Igarashi as the ASJ Co-chairman for the meeting. Professor Hiroya Fujisaki became what amounted to the ASJ Foreign Minister.

Professor Fujisaki and I worked out the details of procedure, responsibility, and funding during one of his visits to Hawaii. The ASA and ASJ approved them. They were successful, and they became the basis for other joint meetings.

Technical Program Organizing Committee (TPOM) arrangements had a hidden surprise. George Sutton, our original Technical Program Chairman, was also an active oceanographer. A scheduled ocean trip was delayed—and delayed—until it coincided with the scheduled time for the TPOM.

We were lucky. Tony Embleton and Eileen were planning to stop in Hawaii on return from Australia—at exactly the time scheduled for the TPOM! Tony graciously agreed, on two-week's notice, to extend his visit and take on the job as Technical Program Chairman. The rest is history. Tony's ability to organize and coordinate made this one of the best technical programs ever.

The TPOM was held at the University of Hawaii. We used a new system for allocating session space. We set five 4 ft×8 ft "Canec" boards against a wall, one for each meeting day. Each had spaces for each meeting room down a side, and major time periods along the top. We preallocated all space and times and pinned small, multicolor identifying cards on the boards. These included committee meetings and social functions. The objective was to replace the race to be first to claim time and space with negotiation among Technical Committee (TC) representatives for exchange. It worked! Tony made some changes, and the procedure has been adopted by others. I've been told that our Japanese colleagues thought we had always done it this way.

To allow members to enjoy Hawaii during the best part of the day without missing important papers, I asked the TC representatives to keep the time period 12 noon to 2 p.m. clear of sessions. But a "work" mentality governed. The ending times for morning sessions and the starting times for afternoon sessions crept toward each other until little was left. After the TPOM was over, acting as the MC, I shortened the time allowed for both morning and afternoon papers until the morning sessions ended at noon, and the afternoon sessions did not start until 2 p.m. I don't recall any later complaints.

I arranged for the two presidents to have side-by-side suites. These were premier ocean-view, two-room suites. At that time, Presidents' Receptions were held in their suites. The idea was that the two presidents could create the environments they wanted, and that ASA and ASJ visitors could easily move from one to the other. It was crowded, but it worked! Japanese hospitality was exquisite, and American hospitality was superb. Everyone enjoyed it—except, possibly, the presidents' wives.

The Accompanying Persons Program was a great success. Frances Burgess and Milly Darby arranged genuine Hawaiian activities for each morning, including history, crafts, food preparation, singers, and dancers. These were not the usual tourist tinsel. Milly also provided the large number of leis that we used lavishly throughout the meeting. For many years afterwards, ladies would tell me at ASA meetings how much they enjoyed what Frances and Milly had done for them.

For this special occasion, I decided to depart from the standard ASA procedure of having two socials, and to replace one with an old-time, traditional banquet. The banquet was so popular that several people actually tried to crash it. The major features were a grand, multicourse Chinese dinner,

[a]Home address: 1269 Mokulua Dr., Kailua, HI 96734. Electronic mail: jcb@hawaii.edu

and Hawaiian entertainment. When attendees reached their tables, they were greeted with bottles of cooled champagne. The loud popping of corks set a great tone!

A pleasant surprise followed the ASA awards ceremony when ASJ President Nimura presented unique awards (with silver medals!) to ASA President James Flanagan and Meeting Chairman John Burgess.

The ASJ created a special "preparation committee" for this meeting. I was greatly surprised at how few of its ASA members were not Fellows. Professor Igarashi agreed to coordinate nominations. A special feature of the banquet was honoring seven new Japanese ASA Fellows.

This history would not be complete without recognizing George Curtis as the Facilities and Equipment Chairman and Ron Darby as the Registration Chairman. George hired university students as projection equipment operators, giving them first-hand experience with a major international technical meeting. Ron's innovative and efficient registration process had the important effect of giving attendees a pleasant first impression.

A final anecdote concerns Bob Beyer, the ASA Treasurer. During the meeting, in the lobby of the hotel, the ASJ casually handed him a large stack of $100 bills. With his attention thoroughly concentrated, he carried this cash across busy Kalakaua Avenue to the Bank of Hawaii—and has never fully recovered from the experience.

The First Joint Meeting was almost the largest ASA meeting and had the largest number of papers presented to that date. It was so popular that it was followed ten years later by a Second Joint Meeting, then after eight years by a Third, and now by a Fourth. Our two Societies were changed.

Anomalous negative dispersion in bone can result from the interference of fast and slow waves

Karen R. Marutyan, Mark R. Holland, and James G. Miller
Laboratory for Ultrasonics, Department of Physics, Washington University in St. Louis,
One Brookings Drive, St. Louis, Missouri 63130
james.g.miller@wustl.edu

Abstract: The goal of this work was to show that the apparent negative dispersion of ultrasonic waves propagating in bone can arise from interference between fast and slow longitudinal modes, each exhibiting positive dispersion. Simulations were carried out using two approaches: one based on the Biot-Johnson model and one independent of that model. Results of the simulations are mutually consistent and appear to account for measurements from many laboratories that report that the phase velocity of ultrasonic waves propagating in cancellous bone decreases with increasing frequency (negative dispersion) in about 90% of specimens but increases with frequency in about 10%.

1. Introduction

Ultrasound studies of bone are often motivated by the goal of developing ultrasonic markers for clinical diagnosis, e.g., for diagnosis of osteoporosis. Cancellous or trabecular bone is a porous material consisting of a solid matrix (trabeculae) and pores filled with fluidlike bone marrow. The mechanical architecture of cancellous bone is anisotropic—bone trabeculae are aligned along the specific directions resulting from stresses placed on bone during its development. Ultrasound propagation in bone has been modeled using Biot theory (Biot, 1955a; 1955b) that was originally introduced in the geophysics literature for application to fluid-saturated porous rocks. Biot theory presents numerous challenges, in part because of the need for values of many parameters of bone that are not readily available in the published literature, even for isotropic media. More recently, Schoenberg theory for stratified media has been applied to model bone as alternating layers of solid and fluid layers (Hughes *et al.*, 1999; Padilla and Laugier, 2000; Wear, 2001). Both the Biot and the stratified model predict the existence of two compressional wave modes, a fast and a slow mode.

Both models enjoy some success in accounting for experimental data. Measurements of the ultrasonic characteristics of bone include the attenuation coefficient, speed of sound, and phase velocity, and their respective frequency dependences. Bone attenuates ultrasound in a manner that is approximately linear with frequency (Chaffai *et al.*, 2000; Wear *et al.*, 2000). The attenuation characteristics are frequently reported as a rate of change with frequency, Broadband Ultrasound Attenuation (BUA) or, if normalized with the propagation distance, normalized Broadband Ultrasound Attenuation (nBUA). Although there is a general consensus on the linear or quasilinear with frequency dependence of attenuation, there is considerable variation in the behavior of the frequency dependence of phase velocity (dispersion) among bone samples. Indeed, measurements from many laboratories indicate that, on average, the phase velocity of ultrasonic waves propagating in cancellous bone decreases with increasing frequency (negative dispersion). However, positive dispersion is observed in approximately 10%–20% of sites measured (Droin *et al.*, 1998; Nicholson *et al.*, 1996; Strelitzki and Evans, 1996; Waters and Hoffmeister, 2005; Wear, 2000). Furthermore, the confusion is aggravated by the fact that the velocity of both the fast and the slow wave varies with the porosity level and

insonation direction relative to the orientation of the trabecular network (the insonation angle). As a consequence, in some experiments both the slow and fast waves are detected and are clearly delineated. However, in a significant number of experiments the time difference between arrivals of the waveforms associated with both waves is small compared to the temporal extent of the ultrasonic pulse. This results in an overlap in the recorded data and makes it impossible to separate fast and slow modes for quantitative assessment.

One goal of the present work is to examine these apparent inconsistencies in measured ultrasonic indices based on causality considerations. For media with an attenuation coefficient that can be approximated as increasing linearly with frequency, such as bone, the nearly local approximations to the Kramers-Kronig relations with one subtraction (Waters *et al.*, 2000) suggest that an increase of velocity with frequency would be expected, which is in sharp contrast with the decrease with frequency that is often reported. If the attenuation coefficient were strictly proportional to frequency, the dispersion would increase logarithmically with frequency. It is on this basis that we refer to the negative dispersion in bone as anomalous. We note that the term "anomalous dispersion" has a different meaning in optics. In this work we propose an alternative explanation for the observed negative dispersion in cancellous bone. Our hypothesis is that the apparent negative dispersion can result from the interference between the fast wave and the slow wave modes that characterize the composite structure of bone.

2. Methods and computer simulations

Computer simulations were carried out using two different models. An input waveform was constructed with the power spectrum peaking at 550 kHz, with the −6-dB spectrum spanning the frequency range of 250 to 850 kHz and the −20-dB spectrum spanning the frequency range of 100 to 1100 kHz, typical of that used for measurements in bone. The temporal extent of the input signal was approximately 3.8 μs. For both simulations, the input signal was propagated through the specimen or through a nonattenuating and nondispersive water path with velocity of 1500 m/s over the identical travel distance to generate sample and reference radio frequency (rf) traces, respectively. Linear plane-wave propagation was assumed. The thickness of bone specimen served as a variable parameter to adjust the overlap of the fast and slow waveforms. The resulting mixed-mode rf traces that correspond to the interference field were analyzed with a phase spectroscopy algorithm previously validated with experimental data (Marutyan *et al.*, 2006; Trousil *et al.*, 2001) to assess the apparent ultrasonic properties of the resulting rf trace.

For one set of simulations, the acoustic propagation was modeled using the modified Biot-Johnson approach following the formalism by Fellah *et al.* (2004) in which the complex Biot coefficients (or "mass" coefficients) were modified in accordance with the dynamic complex tortuosity introduced by Johnson *et al.* (1987). The propagated waveform was obtained by evaluating a complex transmission coefficient $T(\omega, l)$ that depends on frequency ω and specimen thickness l, as well as on bone structural parameters, which were taken to be identical to

TABLE I. Input data for simulations of Biot-Johnson and stratified models.

Parameter	Value
Porosity	0.83
Viscous characteristics length	5 μm
Tortuosity	1.05
Bulk modulus of pore fluid	2.28 GPa
Bulk modulus of elastic solid	20 GPa
Bulk modulus of the bone skeletal frame	3.3 GPa
Shear modulus of the bone skeletal frame	2.6 GPa
Solid bone density	1960 kg/m^3
Fluid density	1000 kg/m^3
Fluid viscosity	10^{-3} kg/(ms)

those for the specimen evaluated in Fellah *et al.* (Table 1). The mathematical expression for the transmission coefficient is defined in Eqs. (24)–(28) of Fellah *et al.* and derived in the appendix of that work.

In a second set of simulations, bone was modeled by independently propagating two modes that correspond to fast and slow waves, both of which were characterized by attenuation and dispersion that were causally linked. The values for nBUA were assigned to be 7 and 20 dB/MHz/cm for the slow and the fast wave, respectively, identical to the values reported in literature (Waters and Hoffmeister, 2005). The corresponding dispersion curves were then calculated from the nearly local Kramers-Kronig approximation. Because the Kramers-Kronig relations are ambiguous to a constant offset, the values of the phase velocity were arbitrarily set to 2100 and 1500 m/s at 300 kHz for the fast and the slow velocity, respectively. Other choices for phase velocities were explored and were shown not to alter significantly the outcome of the simulations, suggesting that the results obtained would be valid for a range of anatomical locations and degrees of porosity. Once the ultrasonic parameters for sound propagation in bone were established, the frequency domain representation of the output waveform was obtained in the following manner:

$$\text{Output}(\omega) = w \cdot \text{Input}(\omega) \cdot H_{\text{fast}}(\omega, l) + (1 - w) \cdot \text{Input}(\omega) \cdot H_{\text{slow}}(\omega, l), \tag{1}$$

where w represents a weighting parameter that varies from 0 to l, $\text{Input}(\omega)$ is the frequency spectrum of the input waveform. The transfer functions $H_{\text{fast}}(\omega, l)$ and $H_{\text{slow}}(\omega, l)$ for the fast and slow modes are given by
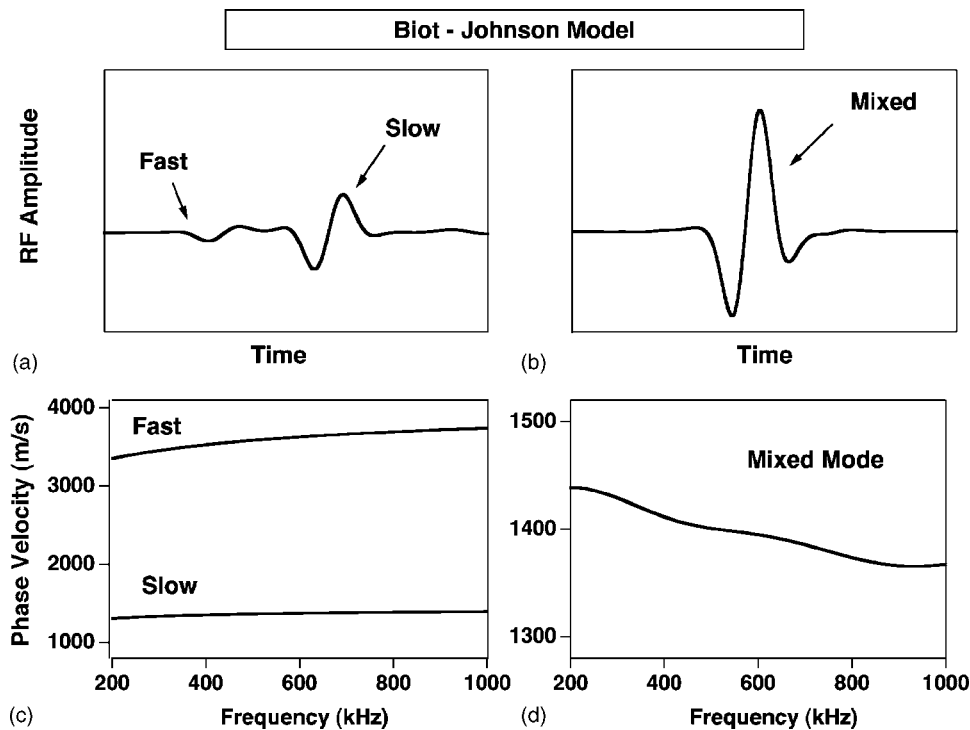
FIG. 1. Simulations for the Biot-Johnson model. Bone parameters used in the simulations are listed in Table I. (a) Radio frequency signal that contains distinctly separable fast and slow waves; (b) radio frequency signal of the mixed mode; (c) theoretical values for phase velocity of the fast and slow wave; (d) phase velocity for the mixed mode ($\Delta t = 0.7 \ \mu s$).
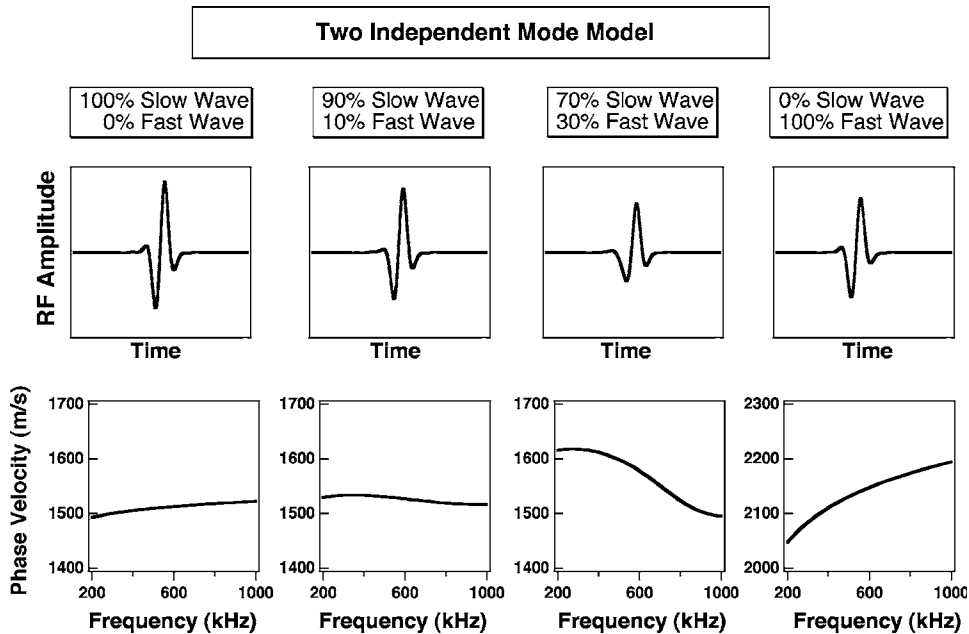
Fig. 2. Results for the two independent mode simulations with the weighting parameter, *w*, of values of 0.0, 0.1, 0.3, and 1.0 from the leftmost to rightmost column, respectively. For all cases Δt is set to 0.6 μs.

$$H_{\text{fast/slow}}(\omega, l) = \exp(- \alpha_{\text{fast/slow}}(\omega) \cdot l) \cdot \exp(i\omega l / c_{\text{fast/slow}}(\omega)), \tag{2}$$

where $\alpha_{\text{fast/slow}}(\omega)$ denotes the (linear with frequency) attenuation coefficient and $c_{\text{fast/slow}}(\omega)$ denotes the (logarithmically increasing) phase velocity for the fast and the slow wave, respectively.

3. Results and discussion

Results from the simulations based on the Biot-Johnson model are shown in Fig. 1. For the given sets of parameters, the phase velocity of the fast and the slow wave was characterized with a theoretical dispersion that is positive and logarithmiclike as seen in panel (c). The midband value of phase velocity for the fast wave lay in the range of 3000 to 4000 m/s, whereas the phase velocities of the slow wave were about 1400 m/s. To characterize the temporal overlap of the two modes, we introduced the parameter Δt, representing the time difference in the arrival times of the fast and slow waves. As the value of Δt decreases, significantly overlapped composite waveforms are produced. For a representative case with $\Delta t \approx 0.7$ μs, a *negative* dispersion was predicted in accordance with the proposed hypothesis [panels (b) and (d) of Fig. 1]. However, the appearance of the composite rf waveform reveals very little, if any, interference effects. Nevertheless, the interference overlap resulted in an apparent negative dispersion, which is substantial (about −70 m/s from 200 kHz to 1 MHz) and could be approximated as quasilinear. This value for the magnitude of dispersion is comparable to values measured experimentally (Droin *et al.*, 1998).

      Results of the second set of simulations are illustrated in Fig. 2, in which composite rf waveforms obtained by superposition of the fast and slow waves are plotted along with the estimated dispersion. The content of the mixed waveform was varied to produce fields ranging from 100% slow wave to 100% fast wave (the leftmost and rightmost columns of Fig. 2, respectively). The second and third columns represent signals that correspond to propagation of the ultrasonic pulse with weighting factor *w* equal to 0.1 and 0.3 (10% and 30% contribution of fast wave), respectively. The overall velocity level of the vertical axes for the dispersion plots is
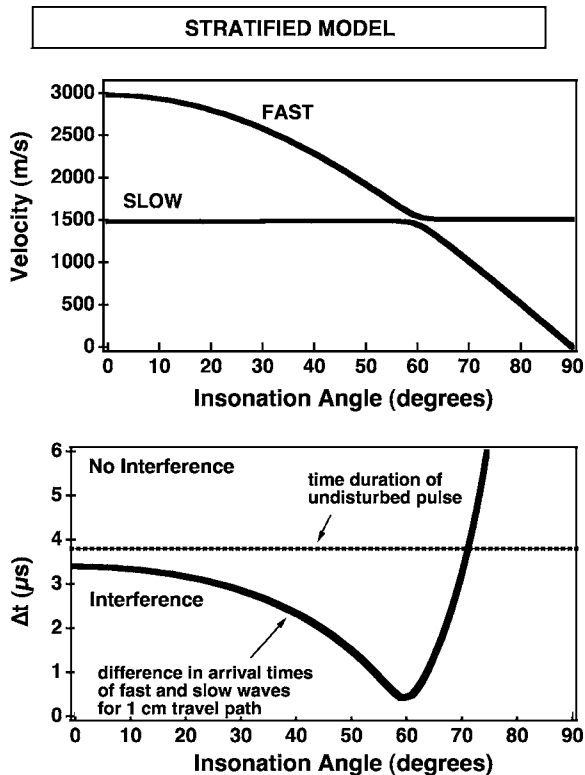
FIG. 3. Predictions from the stratified model of bone: (top panel) anisotropy in velocity of fast and slow waves; (bottom panel) the time difference between arrival time of the fast and slow wave Δt (normalized per propagation distance in cm) plotted vs insonation angle relative to trabecular alignment. The region below the dashed line, which is defined by the pulse length (3.8 μs for current work), is where the interference of fast and slow wave modes occurs.

varied, but the relative scale is preserved so that the phase velocity axis spans a range of 300 m/s in each plot. As expected, for the cases with the existence of only one of the modes (100% slow or 100% fast waves) the ultrasonic parameters (i.e., a logarithmically increasing dispersion and linear attenuation) that served as input for simulations were recovered. For the mixed-mode cases, however, the simulations resulted in an apparent negative dispersion (second and third columns in Fig. 2). Once again a superficial examination of the rf traces reveals no visual indications of the effects due to the mode mixing. For this simulation the temporal overlap was characterized using values of $\Delta t \approx 0.6$ μs. Further investigation revealed that dispersion for the most part remained negative as Δt was varied from 0.4 to 1.0 μs. These values are similar to those used in simulations based on the Biot-Johnson model. In addition, simulations were performed by incrementally increasing the overall level of the velocity of the fast wave $V_{\text{Fast at 300 kHz}}$ from 1900 to 2500 m/s. These simulations (not shown here) similarly revealed mostly negative dispersion for the cases in which $\Delta t < 1$ μs.

In the laboratory setting, the interference of fast and slow waves has been observed in a number of studies. In an early study of bone anisotropy Hosokawa and Otani (1998) observed that "fast and slow waveforms widely overlap in time at incident angles $\theta_I = 20°$ and 30°" and higher. More recently, Lee *et al.* (2003) stated that "since the fast and slow waves were completely overlapped, only one longitudinal wave was observed for propagation in the ML directions." Padilla and Laugier (2000) independently proposed mode mixing to explain the disappearance of "slow wave at high angle of the trabecular alignment." They invoked a stratified model to estimate the arrival time for both modes, and showed that fast and slow waves are

strongly overlapped. In that theoretical work, a study with varying oblique incidence angle θ_I was assumed, and arrival times for slow and fast waves were evaluated as the function of the refraction angle θ_r. For the identical incident angle θ_I, slow and fast waves propagate at *different* refraction angles. Therefore, in our studies, we have chosen to evaluate the difference in arrival time Δt as a function of insonation angle, as illustrated in Fig. 3. The top panel of this figure depicts velocity of the fast and slow waves, and the bottom panel shows the difference in arrival time Δt for a 1-cm ultrasonic path plotted from parallel (0°) to perpendicular (90°) insonation relative to trabecular alignment. Of course, the curve of Δt would scale with the distance, whereas the pulse length would not. The order of Δt at which interference occurs is in the same range at which negative dispersion was shown to occur in our simulations, especially in the range of angles from 50° to 70°. The large values of Δt near perpendicular insonation result from stratified model predictions of nearly zero velocity for the slow wave at that angle. The calculations are based on the bone parameters given in Table 1.

Only phase distortions due to interference effects in the ultrasonic field are considered in the current study. In practice, there are additional complexities, such as phase cancellation effects at the face of a phase-sensitive (piezoelectric) receiver or receiving array, diffraction effects, and potential frequency-dependent reflection and transmission coefficients at the bone boundaries. These effects are usually specific to the experimental setup.

4. Conclusion

Results presented suggest that an interference effect between fast and slow wave modes can be a source for the measurement of an apparent negative dispersion in bone. These simulations using two independent approaches indicate that an apparent negative dispersion can arise even though the individual dispersions for the both fast and slow waves are positive. The degree of interference between fast and slow waves was consistent between both types of simulations and appeared to be comparable to that seen in representative experimental studies.

Acknowledgments

References and links

Biot, M. A. (**1955a**). "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178.

Biot, M. A. (**1955b**). "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191.

Chaffai, S., Padilla, F., Berger, G., and Laugier, P. (**2000**). "*In vitro* measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz," J. Acoust. Soc. Am. **108**, 1281–1289.

Droin, P., Berger, G., and Laugier, P. (**1998**). "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592.

Fellah, Z. E. A., Chapelon, J. Y., Berger, S., Lauriks, W., and Depollier, C. (**2004**). "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73.

Hosokawa, A., and Otani, T. (**1998**). "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722.

Hughes, E. R., Leighton, T. G., Petley, G. W., and White, P. R. (**1999**). "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 811–821.

Johnson, D. L., Koplik, J., and Dashen, R. (**1987**). "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402.

Lee, K. I., Roh, H. S., and Yoon, S. W. (**2003**). "Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot-Attenborough model," J. Acoust. Soc. Am. **144**, 2284–2293.

Marutyan, K. R., Yang, M., Baldwin, S. L., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2006**). "The frequency dependence of ultrasonic velocity and the anisotropy of dispersion in both freshly excised and formalin-fixed myocardium," Ultrasound Med. Biol. **32**, 603–610.

Nicholson, P. H. F., Lowet, G., Langton, C. M., Dequeker, J., and Van der Perre, G. (**1996**). "A comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurement in trabecular bone," Phys. Med. Biol. **41**, 2421–2435.

Padilla, F., and Laugier, P. (**2000**). "Phase and group velocities of fast and slow compressional waves in trabecular bone," J. Acoust. Soc. Am. **108**, 1949–1952.

Strelitzki, R., and Evans, J. A. (**1996**). "On the measurements of the velocity of ultrasound in the os calcis using

short pulses," Eur. J. Ultrasound **4**, 205–213.

Trousil, R. L., Waters, K. R., and Miller, J. G. (**2001**). "Experimental validation of the use of Kramers-Kronig relations to eliminate the phase sheet ambiguity in broadband phase spectroscopy," J. Acoust. Soc. Am. **109**, 2236–2244.

Waters, K. R., and Hoffmeister, B. K. (**2005**). "Kramers-Kronig analysis of attenuation and dispersion in trabecular bone," J. Acoust. Soc. Am. **118**, 3912–3920.

Waters, K. R., Hughes, M. S., Mobley, J., Brandenburger, G. H., and Miller, J. G. (**2000**). "On the applicability of Kramers-Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563.

Wear, K. A. (**2000**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.

Wear, K. A. (**2001**). "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083.

Wear, K. A., Stuber, A. P., and Reynolds, J. C. (**2000**). "Relationships of ultrasonic backscatter with ultrasonic attenuation, sound speed and bone mineral density in human calcaneus," Ultrasound Med. Biol. **26**, 1311–1316.

ACOUSTICAL NEWS—USA

Elaine Moran

Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502

Editor's Note: *Readers of this journal are encouraged to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news items and notices are 2 months prior to publication.*

New Fellows of the Acoustical Society of America

Sergio Beristain
For contributions to international
and national society relationship

David R. Palmer
For contributions to sound
propagation in the ocean

Mark V. Trevorrow
For contributions to high-frequency
underwater acoustics

Michael Vorländer
For contributions to standards and
modeling in architectural acoustics

Report of the 27th F. V. Hunt Postdoctoral Fellow (2005-2006)

Catherine Berchok, Marine Physical Laboratory/Scripps Institution of Oceanography, University of California, San Diego, California 92106

I was fortunate to spend my year as the 2005–2006 F.V. Hunt Post-doctoral Fellow under the guidance of two mentors: Dr. Gerald D'Spain and Dr. John Hildebrand. Because their current research focuses on acoustic localization and bioacoustics, I was able to work on my research topic, studying the Lombard effect in fin whales, from a well-balanced perspective. I thank them, the Hunt family, and the Acoustical Society of America for this amazing learning experience.

The Lombard effect is a subconscious reflex in which an individual adjusts its vocal amplitude in response to perceived changes in the signal-to-noise ratio. Anyone who has ever tried to carry on a conversation in a crowded room has experienced this effect, but it has also been shown to exist in a variety of other animals including birds, primates, frogs, cats, and odontocete whales. Despite this diversity, there are several common trends found across all species: the Lombard effect shows high inter-individual variability, it is highly context dependent, the strongest response is seen for noise sources that fall within the frequency range of the vocalizations, and the increase in signal level for a given change in noise level is relatively small.

It was this last trend that guided much of my efforts over the past year. Because the Lombard effect is fairly small, accurate source localizations are vital to obtaining source levels with small enough error bounds that the actual Lombard effect is not obscured. To this end I developed a method to visualize and quantify the impact of measurement uncertainties on source

localization errors. I then used this method to examine the effects of array design as well as measurement of different properties of the received acoustic field (including time-of-arrival differences, amplitude ratios, and vector sensor bearings) on the reduction of these source location error bounds. I will be presenting these results at the upcoming Acoustical Society meeting this fall and will be submitting a paper for peer review in the near future. Improvements were made in my analysis routines (signal detector, cross correlator, and localizer) based on the findings of this study. For example, I added the relative-relocation cross-correlation technique (an approach developed in the seismic signal processing community and suggested by Dr. Mark McDonald and Dr. John Hildebrand) to improve time-of-arrival difference measurement uncertainty.

It is clear that when using an array of widely separated acoustic sensors to measure the Lombard effect, the optimal situation exists when the calling animal is within the perimeter of the array, and the noise source is either homogeneous (like wind-generated noise) or at a range significantly greater than the dimensions of the array (in the case of a point source such as a transiting ship). In this case the localization errors are minimized and the noise field received by the animal can be determined fairly easily. For this reason a widely spaced acoustic sensor array that can be redeployed frequently as the animals move about is the ideal solution. As a first stage in the design of this array, I collaborated with Richard Sears (Mingan Island Cetacean Study, Québec, Canada) to design and build a single autonomous recording buoy which is currently being used to record whales in the field. I am also working under the guidance of Bertel Møhl and his colleagues at Aarhus University in Denmark to incorporate their GPS-based time synchronization module into the buoy, which will greatly improve the localization accuracy of the array.

The high inter-individual variation was another aspect of the Lombard effect that caught our interest over the past year. Working with Dr. Ann Bowles at the Hubbs-SeaWorld Research Institute and with Dr. Pat Moore, Dr. Jim Finneran, and Dr. Dorian Houser at the SPAWAR Systems Center in San Diego, we developed an experimental protocol for measuring the extent of and possible reasons for this variability between individuals in a controlled environment. Presently, it is in review by the Hubbs-SeaWorld Research Institute's Institutional Animal Care and Use Committee.

I would once again like to extend my deepest thanks to the Hunt family for providing me with this opportunity. The questions raised during the research funded by this fellowship have led me into many new interesting areas of research that I had not been familiar with before this year, and I am fortunate to have met so many knowledgeable and helpful researchers in these areas. We hope to continue this work in the future and have submitted a detailed, multi-faceted proposal to the Office of Naval Research to further pursue various aspects of the Lombard effect in marine animals.

List of Publications from the past year:

Berchok, C. L., D'Spain, G. L., and Hildebrand, J. A. (**2006**) "Reducing source localization errors: a visualization method to help guide the design, calibration and use of widely separated acoustic sensor arrays," *152nd Meeting of the Acoustical Society of America Joint Meeting With* the Acoustical Society of Japan, Honolulu, Hawaii, 28 Nov. 2 Dec.

Berchok, C. L., D'Spain, G. L., and Hildebrand, J. A., "Comments on marine mammal localization errors using an array of widely separated sensors" (*submission pending*).

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future. The month/year notation refers to the issue in which a complete meeting announcement appeared.

2006

28 Nov. 2 Dec.	152nd Meeting of the Acoustical Society of America joint with the Acoustical Society of Japan, Honolulu, Hawaii [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; www: http://asa.aip.org]. Deadline for receipt of abstracts: 30 June 2006.

2007

4–8 June	153rd Meeting of the Acoustical Society of America, Salt Lake City, Utah [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; www: http://asa.aip.org].
22–24 Oct.	NoiseCon 2007, Reno, NV [Website is: www.inceusa.org/nc07]
27 Nov.– 2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, Louisiana (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; www: http://asa.aip.org].

2008

28 July–1 Aug.	9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton, CT 06340-1609, Tel. 860-572-0680; Web: www.icben.org. E-mail icben2008@att.net].

Cumulative Indexes to the Journal of the Acoustical Society of America

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below.

Volumes 1–10, 1929–1938: JASA, and Contemporary Literature, 1937–1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; Nonmembers $10.
Volumes 11–20, 1939–1948: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 395. Out of Print.
Volumes 21–30, 1949–1958: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75.
Volumes 31–35, 1959–1963: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90.
Volumes 36–44, 1964–1968: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print.
Volumes 36–44, 1964–1968: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print.
Volumes 45–54, 1969–1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 55–64, 1974–1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 65–74, 1979–1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound).
Volumes 75–84, 1984–1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
Volumes 95–104, 1994–1998: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632, Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound).
Volumes 105–114, 1999–2003: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 616, Price: ASA members $50; Nonmembers $90 (paperbound).

ACOUSTICAL STANDARDS NEWS

Susan B. Blaeser, Standards Manager

ASA Standards Secretariat, Acoustical Society of America, 35 Pinelawn Rd., Suite 114E, Melville, NY 11747 [Tel.: (631) 390-0215; Fax: (631) 390-0217; e-mail: asastds@aip.org]

George S. K. Wong

Acoustical Standards, Institute for National Measurement Standards, National Research Council, Ottawa, Ontario K1A 0R6, Canada [Tel.: (613) 993-6159; Fax: (613) 990-8765; e-mail: george.wong@nrc.ca]

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes Catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please, contact Susan B. Blaeser.

Comments are welcomed on all material in Acoustical Standards News.

This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: http://asa.aip.org

Standards Available for Use in the Classroom

In many technical areas it is essential that students become familiar with standards while they are still in school. Until recently, however, the high cost of standards was a barrier to using them in the classroom. Under a special ASA program, teachers who are ASA members can obtain a site license to reproduce standards needed in the classroom at a low cost.

Nearly 120 American National Standards developed by Accredited Standards Committee S1 Acoustics, S2 Mechanical Vibration and Shock, S3 Bioacoustics, and S12 Noise, are available under this program.

For more information contact the ASA Standards Secretariat at 631 390-0215. The complete list of American National Standards developed by these committees can be seen at the ASA Standards Store at http://asastore.aip.org/ or request a catalog from the Secretariat.

New Working Groups in S12

Accredited Standards Committee S12, Noise, recently approved the formation of two new working groups. S12/WG 45 "Measurement of Occupational Noise Exposure from Telephone Equipment" is chaired by A. K. Woo. The scope of S12/WG45 is: "Development of standard method(s) for measuring occupational noise exposure from telephone handsets, headsets, and headphones. This work may result in either the revision of ANSI S12.19-1996 or the development of a new standard."

S12/WG 46 "Acoustical Performance Criteria for Relocatable Classrooms" is co-chaired by T. Hardiman and P. D. Schomer. The scope of S12/WG 46 is: "To develop a standard that focuses on the challenges and unique circumstances surrounding the use and relocation of relocatable classrooms with a focus on labeling and methods to quantify acoustical performance."

WG membership is open to all interested experts and there is no fee to participate in a working group. If you are interested in more information about these or other standards working groups, please contact the Secretariat.

Standards Meetings Calendar National

The 152nd ASA Meeting, Honolulu, Hawaii, at the Sheraton Waikiki Hotel, 28 November–2 December 2006, the ASA Committee on Standards (ASACOS) and ASACOS STEERING Committees will meet as below:

•*Tuesday, 28 November 2006*

ASACOS Steering Committee

•*Wednesday, 29 November 2006*

ASA Committee on Standards (ASACOS). Meeting of the Committee that directs the Standards Program of the Acoustical Society.

International Standards Committee Meetings

•*16–20 April 2007*

Meeting of **ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines**, Prague, Czech Republic

SC 5 and all of its working groups will be joined by the following additional working groups:

TC 108/WG 31 Balancing

TC 108/SC 2/WG 1 Vibration of machines

TC 108/SC 2/WG 7 Vibration of machines with active magnetic bearings

TC 108/SC 2/WG 10 Basic techniques for vibration diagnostics

•*23–27 April 2007*

Meetings of **ISO/TC 108, TC/108 SC 2, TC 108/SC 3 and TC 108/SC 6**, joined by their working groups (except those listed above), will meet in Berlin, Germany.

Recent International Standards Meetings

September and October 2006 were busy months for meetings of the International Standards Committees for which ASA administers the U.S. Technical Advisory Groups, as well as other international committees that are of interest to ASA's standards committees. The meetings began in September in Barcelona Spain, where **IEC/TC29 Electroacoustics** and its working groups met.

During the following week Madrid hosted the meetings of **ISO/TC 43 Acoustics; TC 43/SC 1 Noise**; and **TC 43/SC 2 Building Acoustics**, and their working groups.

During the same week, the **Consultative Committee on Acoustics, Ultrasound and Vibration** (CCAUV) met in Sèvres, France.

Members of **ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock** and its European counterpart, **CEN/TC 231**, and their working groups met in Bratislava, Slovakia in October.

Results from these meetings were not available by press time. Items of

particular interest will appear in future issues. For information about these Committees or any other aspects of the ASA Standards Program, please contact the Secretariat.

Standards News from the United States

(Partially derived from *ANSI Reporter* and *ANSI Standards Action*, with appreciation)

American National Standards Call for Comment on Proposals Listed

This section solicits comments on proposed new American National Standards and on proposals to revise, reaffirm, or withdraw approval of existing standards. The dates listed in parenthesis are for information only.

ASA (ASC S1) (Acoustical Society of America)
Revisions

BSR S1.40-200x, Specification and Verification Procedures for Sound Calibrators [revision of ANSI S1.40-1984 (R2001)]

The standard specifies performance requirements for the sound pressure level, frequency, and total distortion generated by a sound calibrator. It also provides requirements for the influence of environmental conditions, for electromagnetic compatibility, and for instrument marking and documentation. It gives details of the tests necessary to verify that a model of sound calibrator conforms to all the requirements, as well as details of the method for periodic testing of a sound calibrator. (28 August 2006)

ASTM (ASTM International)

BSR/ASTM Z2340Z (F2544)-200x, Test Method for Determining a Weighted Sound Power Level of Central Vacuum Power Units (new standard) (3 September 2006)

CEA (Consumer Electronics Association)
New Standards

BSR/CEA 2031-200x, Testing and Measurement Methods for Mobile Loudspeaker Systems (new standard)

Defines test procedures for rating the performance and physical size of mobile loudspeakers, and requirements for reporting these characteristics. CEA 2031, when used in conjunction with CEA 2006-A, Testing & Measurement Methods for Mobile Audio Amplifiers, enables consumers to select mobile loudspeakers with power handling capabilities that are appropriate for the power output characteristics of their mobile amplifiers. (11 September 2006)

SCTE (Society of Cable Telecommunications Engineers)
New Standards

BSR/SCTE 24-21-200x, BV16 Speech Codec Specification for Voice Over IP Applications in Cable Telephony (new standard)

This document contains the description of the BV16 speech codec. BV16 compresses 8-kHz sampled narrowband speech to a bit rate of 16 kb/s by employing a speech coding algorithm called two-stage noise feedback coding (TSNFC).

ATIS (Alliance for Telecommunications Industry Solutions)

BSR T1.421-2001 (R200x), In-Line Filter for Use with Voiceband Terminal Equipment Operating on the Same Wire Pair with High Frequency (up to 12 MHz) Devices (reaffirmation of ANSI T1.421-2001)

This standard addresses the steady-state operation of both plain old telephone service (POTS) and digital subscriber line (DSL) services. Various tests and requirements are described to minimize mutual interactions in terms of continuous noise or signal degradation. Transient events such as impulsive noise and large current spikes introduced by POTS signaling or DSL state changes have not been considered and are left for further study.

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards. The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

EIA (Electronic Industries Alliance)

BSR/EIA 364-45A-2000 (R200x), Firewall Flame Test Procedure for Electrical Connectors (reaffirmation of ANSI/EIA 364-45A-2000)

Establishes test methods to determine the ability of a mated electrical firewall connector to resist specified flame and vibration during a 20-minute exposure. (26 September 2006)

ARI (Air-Conditioning and Refrigeration Institute)
New Standards

BSR/ARI 370-200x, Sound Rating of Large Outdoor Refrigerating and Air-Conditioning Equipment (new standard)

This standard applies to the outdoor portions of factory-made commercial and industrial large outdoor refrigerating and air-conditioning equipment, including heat pumps, used for refrigerating or air-conditioning of spaces. (24 October 2006)

Final Actions on American National Standards
The standards actions listed below have been approved by the ANSI Board of Standards Review (BSR) or by an ANSI-Audited Designator, as applicable.

ASA (ASC S2)
Withdrawals

ANSI S2.47-1990, Vibration of Buildings—Guidelines for the Measurement of Vibrations and Evaluation of Their Effects on Buildings [withdrawal of ANSI S2.47-1990 (R2001)] (19 May 2006)

IEEE (Institute of Electrical and Electronics Engineers)

BSR/IEEE C37.082-1982 (R200x), Standard Methods for the Measurement of Sound Pressure Levels of AC Power Circuit Breakers [reaffirmation of ANSI/IEEE C37.082-1982 (R2000)]

Establishes uniform guidelines for measurement and reporting of sound produced by AC power circuit breakers. Intended for use in measurement of the sound produced by outdoor circuit breakers in a free-field environment. Methods may be used indoors or in a restricted field, provided that precautions are observed in measurement and interpretation of results. Three types of tests are described: design tests, conformance tests, and field tests.

New National Adoptions

AMCA (Air Movement and Control Association)
New Standards

ANSI/AMCA 301-2006, Methods for Calculating Fan Sound Ratings from Laboratory Test Data (new standard): 8/3/2006

UL (Underwriters Laboratories, Inc.)
Revisions

BSR/UL 299 CAN/ULC-S504-200x, Standard for Safety for Dry Chemical Fire Extinguishers (revision of ANSI/UL 299/ULC-S504-2002) This proposal bulletin includes revisions to: correct a reference; operation test requirements; vibration test requirements; and temperature and humidity test requirements for the nameplate exposure and adhesion tests. (25 September 2006)

Information Concerning:
U.S. National Committee of the IEC

USNC TAG to IEC/TC 113—Nanotechnology Standardization for Electrical and Electronic Products and Systems

Call for Participants

The Technical Management Committee (TMC) of the U.S. National Committee for IEC has assigned the National Electrical Manufacturers' Association (NEMA) as the TAG Administrator for the newly established IEC/TC 113—Nanotechnology standardization for electrical and electronic products and systems. The interim scope of the TC is as follows:

Scope:

Standardization in the field of nanotechnology relevant to electricity and related technologies within the area of the IEC. These are especially electronics, magnetics and electromagnetics, electroacoustics, multimedia, telecommunication, and energy production. Specific topics are terminology and symbols, measurement and performance, reliability, design and development, electromagnetic compatibility, safety and environment. NOTE: Liaison will be ensured with other technical committees inside the IEC; relevant national, regional and international standardization bodies and organizations; as well as with other interested organizations and networks world-wide. Special attention will be given to the Liaison with the ISO TC 229 "Nanotechnology" to ensure coordination and to prevent duplication of work. Ken Gettman of NEMA is now in the process of organizing this TAG and anyone interested in participating is invited to contact him as follows: Kenneth E. Gettman, NEMA, 1300 North 17th Street, Suite 1752, Rosslyn, VA 22209; PHONE: (703) 841-3254; FAX: (703) 841-3354; e-mail: ken_gettman@nema.org.

Newly Published ISO and IEC Standards
Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization.

ISO Standards

ACOUSTICS (TC 43)

ISO 140-16:2006, Acoustics—Measurement of sound insulation in buildings and of building elements—Part 16: Laboratory measurement of the sound reduction index improvement by additional lining

ISO 717-1/Amd1:2006, Acoustics—Rating of sound insulation in buildings and of building elements—Part 1: Airborne sound insulation in buildings and of interior building elements—Amendment 1: Rounding rules related to single number ratings and single number quantities

ISO 717-2/Amd1:2006, Acoustics—Rating of sound insulation in buildings and of building elements—Part 2: Impact sound insulation—Amendment 1

ISO 17201-2:2006, Acoustics—Noise from shooting ranges—Part 2: Estimation of muzzle blast and projectile sound by calculation

ISO 16832:2006, Acoustics—Loudness scaling by means of categories

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 16063-15:2006, Methods for the calibration of vibration and shock transducers—Part 15: Primary angular vibration calibration by laser interferometry

TRACTORS AND MACHINERY FOR AGRICULTURE AND FORESTRY (TC 23)

ISO 22867/Cor1:2006, Forestry machinery—Vibration test code for portable hand-held machines with internal combustion engine—Vibration at the handles—Corrigendum

SIEVES, SIEVING AND OTHER SIZING METHODS (TC 24)

ISO 20998-1:2006, Measurement and characterization of particles by acoustic methods—Part 1: Concepts and procedures in ultrasonic attenuation spectroscopy

IEC Standards

SEMICONDUCTOR DEVICES (TC 47)

IEC 60749-35 Ed. 1.0 b:2006, Semiconductor devices—Mechanical and climatic test methods—Part 35: Acoustic microscopy for plastic encapsulated electronic components

ISO Draft Standard

ACOUSTICS (TC 43)

ISO/DIS 13473-5, Characterization of pavement texture by use of surface profiles—Part 5: Determination of megatexture (2 November 2006)

IEC Draft Standard

77/319/FDIS, IEC 61000-4-1 Ed.3: Electromagnetic compatibility (EMC)—Part 4-1: Testing and measurement techniques—Overview of IEC 61000-4 series (1 September 2006)

29/604/FDIS, IEC 60118-4 Ed.2: Electroacoustics—Hearing aids—Part 4: Induction loop systems for hearing aid purposes—Magnetic field strength (1 September 2006)

87/357/FDIS, IEC 60565 Ed.2: Underwater acoustics-Hydrophones—Calibration in the frequency range 0,01 Hz to 1 MHz (6 October 2006)

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Acoustical Standards News 2391

29/607/FDIS, IEC 61672-3 Ed.1: Electroacoustics—Sound level meters—Part 3: Periodic tests (8 September 2006)

International documents submitted to the U.S. for vote and/or comment

Some of the documents processed recently by the ASA Standards Secretariat. Dates in parentheses are deadlines for submission of comments and recommendation for vote, and they are for information only.

U.S. TAG	ISO and IEC documents
S1	**IEC/FDIS 60118-4 Ed. 2 (29/604/FDIS)** "Electroacoustics—Hearing aids —Part 4: Induction loop system for hearing aid purposes—Magnetic field strength" (14 August 2006)
S12	**Second ISO/CD 9612** "Acoustics—Measurements and calculation of occupational noise exposure—Engineering method" (Revision of ISO 9612:1997) (10 August 2006)
	ISO NWIP 13742-2 and ISO/CD 13742-2 "Acoustics—Measurement of sound absorption properties of road surfaces in situ—Part 2: Spot method for low absorption surfaces" (18 August 2006)
	ISO/CD 1683 "Acoustics—Preferred reference quantities for acoustical levels"

BOOK REVIEWS

P. L. Marston

Physics Department, Washington State University, Pullman, Washington 99164

These reviews of books and other forms of information express the opinions of the individual reviewers and are not necessarily endorsed by the Editorial Board of this Journal.

Editorial Policy: *If there is a negative review, the author of the book will be given a chance to respond to the review in this section of the Journal and the reviewer will be allowed to respond to the author's comments. [See "Book Reviews Editor's note," J. Acoust. Soc. Am.* **81***, 1651 (May 1987).]*

Handbook of Underwater Acoustic Engineering

Heinz G. Urban

The purpose of the book, quoting the author "is to present the principles of underwater engineering in as simple and concise a form as possible." The book was first published in German in 2000 and, as a result of the wide use both in Germany and outside, was rewritten in English, based on an updated German version. It was published by the Naval Systems Division of STN ATLAS Elektronic GmbH. To obtain a copy of the book, the request is sent to the following address: ATLAS Elektronik GmbH, Naval Systems Division, Attn: Prof. Manfred Siegel, Sebaldsbrucker Heerstr. 235, D-28305 Bremen, Germany. The policy of the publisher is not to sell it, but to request a donation (70 Euros or approximately $100.00 US) to a charity organization (German Maritime Rescue Service).

To put the book in context, it is between Urick's *Principles of Underwater Sound (3rd Edition)*, which has an extensive reference base and presents physical data as collected at sea and in laboratory measurements, organized around the terms of the sonar equation and Jensen, Kuperman, Porter, and Schmidt's *Computational Ocean Acoustics*, which is an analytic treatise, necessarily focused on a subset of sonar equation subjects, and provides detail that allows a reader to write computer codes to calculate, e.g., acoustic pressure amplitude as a function of depth and range, leading to transmission loss, a sonar equation parameter.

The text has 11 chapters and six appendices. This reviewer found that it naturally divided into three parts: an introduction to basic units and fundamental acoustic concepts, application of those concepts in the ocean environment or to acoustic devices used in the ocean and discussion of sonar equation terms, use, and assumptions necessary that allow the sonar equation to be used as a design tool. Following the discussion of basic acoustic concepts, reflection and scattering from ocean boundaries are addressed and next propagation. Transduction, transducers, and array directivity are discussed, and then underwater noise. A chapter is devoted to signal processing and introduces signal-to-noise ratios, detection concepts and the book concludes with a chapter that discusses the sonar equation terms, their use in system design, and some specific problems (examples of sonar equation calculations).

It is written in a very "reader friendly" style, near to conversational, and clearly benefits from the author's extensive background in a very complex subject. The value of the text lies in its clear, easy-to-grasp statements of equations, limitations, approximations, most commonly used techniques, etc., that provide a reader with an understanding and capability to work in the world of underwater acoustics. The techniques, formulas, and concepts that have stood "the test of time" are the focus of this text. With the two books noted above, it is a member of a "triad" that should be on the bookshelf of those working in this field.

The bibliography following each chapter could be more complete, but is sufficient to provide the reader with a good start toward the exploring the literature base behind each topic.

DAVID BRADLEY
Applied Research Laboratory
Pennsylvania State University
State College, Pennsylvania 16804-0030

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

The purpose of these acoustical patent reviews is to provide enough information for a Journal reader to decide whether to seek more information from the patent itself. Any opinions expressed here are those of reviewers as individuals and are not legal opinions. Printed copies of United States Patents may be ordered at $3.00 each from the Commissioner of Patents and Trademarks, Washington, DC 20231. Patents are available via the Internet at http://www.uspto.gov.

6,968,740

43.20.Hq SYSTEMS AND METHODS FOR DETERMINING AN ACOUSTIC AND/OR THERMAL CHARACTERISTIC OF A TARGET MATERIAL

Brandon Dillan Tinianov, assignor to Johns Manville International Incorporated
29 November 2005 (Class 73/645); filed 26 January 2004

A number of material-property detectors transmit sound waves through the substance to be tested and measure the resulting spectrum in various ways. The method described in this very short patent would simply measure the resulting signal amplitude, but would also measure the density of the material under test. The intended application is testing insulation material for building construction.—DLR

7,035,166

43.30.Pc 3-D FORWARD LOOKING SONAR WITH FIXED FRAME OF REFERENCE FOR NAVIGATION

Matthew Jason Zimmerman and James Henry Miller, assignors to FarSounder, Incorporated
25 April 2006 (Class 367/88); filed 17 October 2003

A forward-looking (or side-looking, or bottom-looking) sonar system incorporates a transmit transducer for projecting a signal into the water ahead of the vessel and a phased array of receivers to provide return signals to a computer, which then determines azimuthal and elevation angles and times of arrival of the return signals. The system is equipped with roll and tilt sensors, GPS receiver, and compass. The acoustic information plus the roll and tilt information is processed to create a 3-D image of the space ahead of the vessel relative to a fixed frame of reference, i.e., the earth.

Advanced signal processing techniques allow the computer to extract targets from the raw data and other features of the system enable the suppression of multipath targets.—WT

7,049,730

43.35.Ei ACOUSTIC DRIVER ASSEMBLY FOR A SPHERICAL CAVITATION CHAMBER

Ross Alan Tessien et al., assignors to Impulse Devices, Incorporated
23 May 2006 (Class 310/323.12); filed 11 March 2005

Sonoluminescence is an interesting phenomenon that has been studied since the 1930s. Most experimental setups use high-intensity sound waves to stimulate a liquid in a cavitation chamber. This patent (the third in a series of

continuations) describes a balanced, piezoelectric inertia transducer bonded to the surface of a spherical chamber. The arrangement is said to provide more efficient coupling of energy than prior art.—GLA

7,053,529

43.35.Pt TORSIONAL ACOUSTIC WAVE SENSOR

Terence J. Knowles, assignor to TexZec, Incorporated
30 May 2006 (Class 310/333); filed 1 July 2003

This patent discloses the use of torsional waves in a thin substrate to create a touch-activated switch suitable for hazardous environments and other high-value specialty application. It is said that positioning transducer **12** at an appropriate radius and orientation will result in a circulating shear
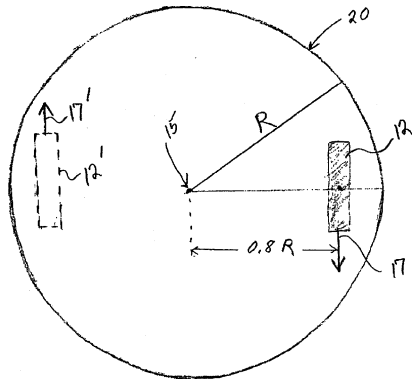
wave traveling around the center of the raised feature **15**. This wave will be strongly damped if the raised area is touched by a finger. This device seems rather complicated and trouble-prone for the task of simply switching something. The patent gives no performance details or detection criteria by which a contact closure might be judged.—JAH

7,052,459

43.35.Yb METHOD AND APPARATUS FOR CONTROLLING ULTRASOUND SYSTEMS

Michael Joseph Washburn et al., assignors to General Electric Company
30 May 2006 (Class 600/437); filed 10 September 2003

In one embodiment, a user interface for controlling an ultrasound system is provided. The user interface includes multiple selectable elements for controlling the ultrasound system and multiple identifiers. Each identifier corresponds to one of the selectable elements and associates control
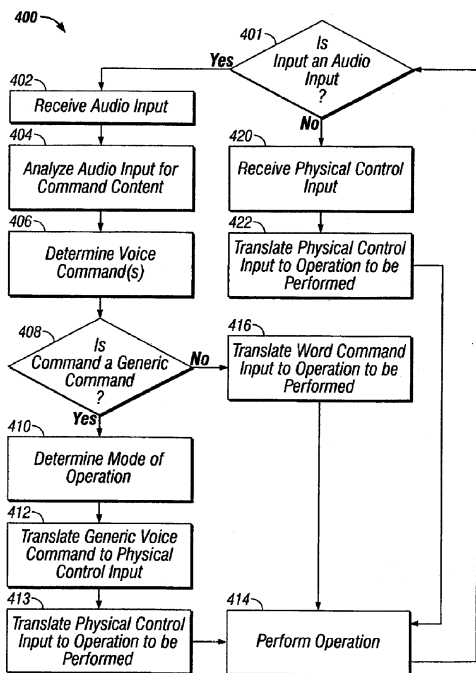
commands with that element. In a second embodiment, the method includes associating a set of identifiers with multiple operations for controlling the ultrasound system, receiving control commands, and performing operations on the basis of receiving control commands corresponding to one or more sets of identifiers.—DRR

7,024,935

43.35.Zc ELECTROMAGNETIC ACOUSTIC TRANSDUCERS

David Paige and Robert Andrew Mercel, assignors to Pii Limited
11 April 2006 (Class 73/643); filed in the United Kingdom 5 February 2002

Assembly **4** is mounted in a "pig" that is sent through an operating gas pipeline **2**. Magnet **30** magnetizes the steel pipeline wall **6**. Ultrasound is generated in the steel pipeline wall by eddy-current excitation via an ac magnetic field **8** produced by an ac current passed through coil **36**. This device is termed as an electromagnetic acoustic transducer (EMAT). Coil **36**
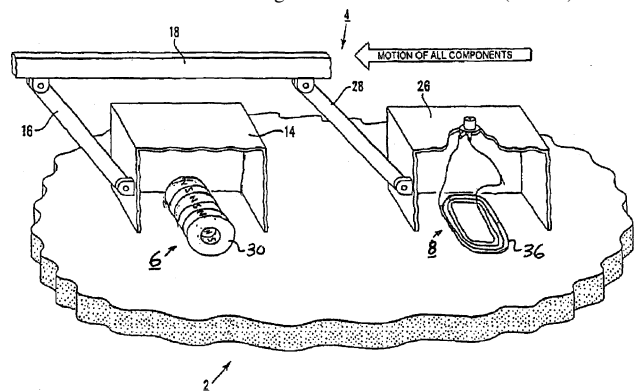
is then switched to a receiver to detect the subsequent ultrasound so generated. The strength of the ultrasound so detected is modulated by nonuniformities in the steel wall. This transducer is asserted to operate over a wider range of operating environments than heretofore. Other magnet and coil arrangements are shown.—AJC

7,040,557

43.35.Zc SYSTEM AND METHOD FOR PULVERIZING AND EXTRACTING MOISTURE

William Graham et al., assignors to Power Technologies Investment Limited
9 May 2006 (Class 241/1); filed 1 April 2004

Sensor **108**, which might be an acoustic sensor in material crusher-dryer **800**, monitors the flow rate of material from blender **102** into inlet tube **12**. An acoustical emission sensor **702** monitors 50–950-kHz ultrasound generated by material flowing through inlet tube **12**, venturi **18**, airflow generator **32**, and housing **35**. The resonant frequency sound
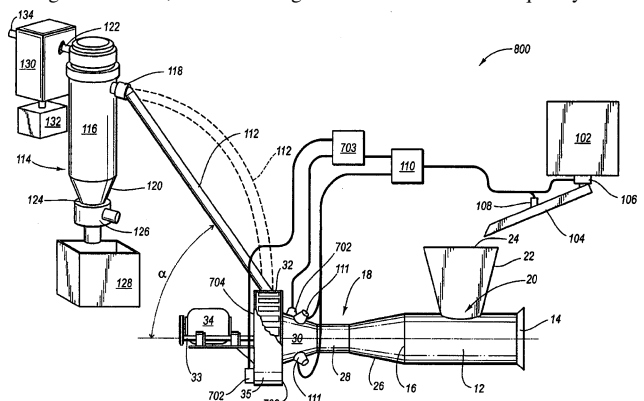
received by **702** indicates the volumetric flow rate. Controller **703** communicates with processor **110** to control that flow rate through flow valves **111** or a diverter (not shown). Sensor **702** also receives sounds that may indicate improper balance or dislodged blades on fan **32**.—AJC

7,044,163

43.35.Zc DRAG REDUCTION IN PIPE FLOW USING MICROBUBBLES AND ACOUSTIC ENERGY

Joline Fan and Zhe Cui, assignors to The Ohio State University
16 May 2006 (Class 137/828); filed 10 February 2005

Gas microbubbles and ultrasound are injected into flowing fluid to reduce the drag of liquid flow in pipes. The author claims a microbubble gas injector **10** comprising a length of pipe **12** with 50- to 400-μm-diam gas injector orifices **14** and a flow restrictor **18**, causing negative pressure to draw the gas into the liquid stream inside the pipe. Drag reductions of up to 30% are cited. Also claimed is the application of acoustic energy into the
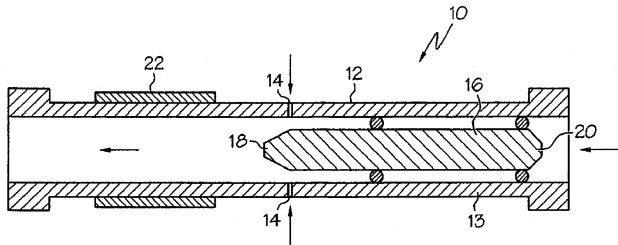
fluid via transducer **22** to cause the microbubbles to move away from the pipe wall. With microbubbles of average diameter of 300 μm, but without acoustic energy, a 125 ml/s flow rate occurred, accompanied by some pressure head fluctuations due to chaotic flow. When up to 600 W (sic) of acoustic energy at 16–20 kHz was applied, the flow rate increased to 130 ml/s, and the pressure fluctuations decreased, indicating organized liquid flow.—AJC

7,053,319

43.35.Zc ELECTRONIC WEIGHING APPARATUS UTILIZING SURFACE ACOUSTIC WAVES USING SENSORS OPERATING AT DIFFERENT FREQUENCIES, HAVING TEMPERATURE COMPENSATION, AND A PUSH OSCILLATOR

Vyacheslav D. Kats and Arnold S. Gordon, assignors to Circuits and Systems, Incorporated
30 May 2006 (Class 177/210 FP); filed 1 February 2002

Weighing apparatus **10** has adjacent surface acoustic wave (SAW) devices **20** and **22** installed on supports **17** and **19** in a cutout **15** within the aluminum weighing beam **14**, whose end will deflect as much as 200 μm downward when a load is placed in pan **16**. SAW devices **20** and **22** are separated by a 20 μm gap **g**, which allows acoustical coupling between said devices. The surface wave induced in **22** from **20** is amplified and fed back to **20**, causing system self-oscillation of a frequency determined by the phase shift of the wave coupled from **20** to **22**. As weight is added, the phase shift is increased, and the frequency reduced accordingly. Several modal
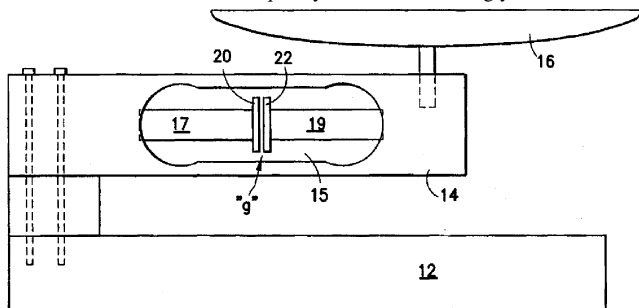
frequencies are possible. The modal frequency found to have a minimal or zero temperature coefficient is assured by driving the system briefly with a "push" oscillator, then measuring the subsequent free oscillating frequency as the weight-indicating value. The claims include the coupling means, the push oscillator application, a second oscillator SAW pair with a weight-dependent self-oscillation frequency, and other arrangements more repeatable under large beam displacements.—AJC

7,053,522

43.35.Zc SURFACE ACOUSTIC WAVE SENSOR

Mauricio Pereira da Cunha, assignor to University of Maine System Board of Trustees
30 May 2006 (Class 310/313 A); filed 26 February 2003

A surface acoustic wave (SAW) delay line **10** is claimed with sensitivity to the properties of liquid **23**, apart from those of pure water. Interdigital transducer (IDT) **16**, **18** on piezoelectric substrate **14** of the lanthanum gallium oxide family (LGX) is cut at selected Euler angles. For a minimum temperature coefficient, Euler angles of ϕ being 0°, θ being about
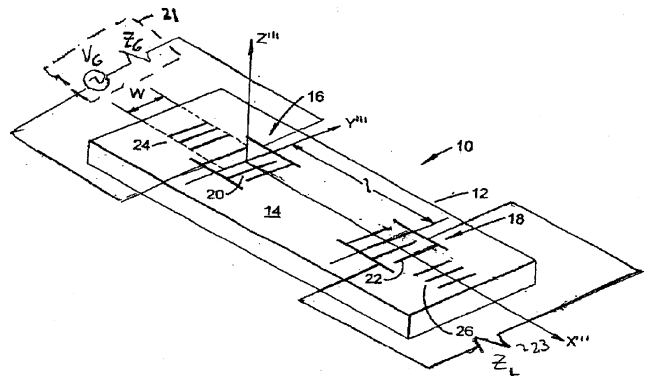
14°, and ψ being about 90° are claimed, along with there being a minimum sensitivity to water, while retaining sensitivity to chemical or biological agents. Changes in a SAW "parameter" (loss? phase?) are said to occur in the presence of any of those agents.—AJC

7,053,524

43.35.Zc SURFACE ACOUSTIC WAVE SENSOR OR IDENTIFICATION DEVICE WITH BIOSENSING CAPABILITY

Peter J. Edmonson *et al.*, assignors to P.J. Edmonson Limited
30 May 2006 (Class 310/313 D); filed 31 May 2005

A radio-frequency identification biosensor **200**, all on a piezoelectric substrate, sends surface waves **240** on that substrate toward reflectors **232**, **234**, and **236**, the last of which, **236**, is coated with a biolayer **235** under a fluid chamber **262–265**. When a biomaterial is present in (flows through)

2397   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2397
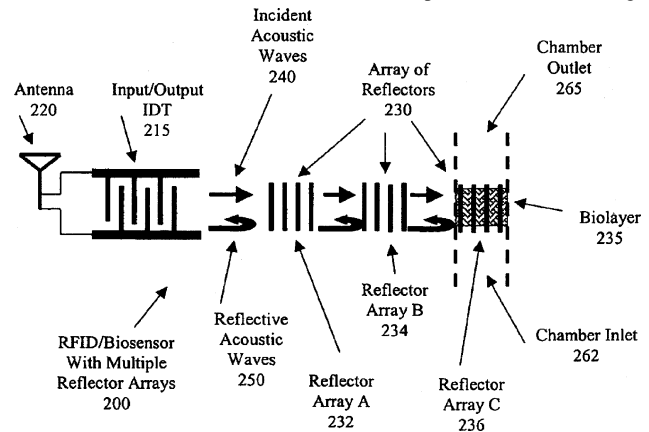
that chamber, biolayer **235** is changed, and the acoustic velocity of the surface wave in reflector **236** is also changed, altering the delay, or phase, or amplitude of the wave reflected by **236**. Detection algorithms for the phase shift of the reflected wave and a "reference modulation code" are claimed.—AJC

7,042,138

43.38.Ar PIEZOELECTRIC ACOUSTIC TRANSDUCER

Yuko Yokoi *et al.*, assignors to Murata Manufacturing Company, Limited
9 May 2006 (Class 310/324); filed in Japan 10 April 2003

This patent describes a shrink-curved piezoelectric laminate consisting of plastic film adhered to a thin piezo sheet. It is argued here that by using upper and lower films of plastics having different hardening temperatures, a curvature may be developed in the parts as they cool and that the curved parts will have a higher radiation efficiency for sound production.—JAH

7,043,035

43.38.Ar MINIATURE MICROPHONE

Zacharias M. Rittersma *et al.*, assignors to SonionMicrotronic Nederland B.V.
9 May 2006 (Class 381/173); filed 7 December 2000

The patent discusses improvements in miniature microphones for
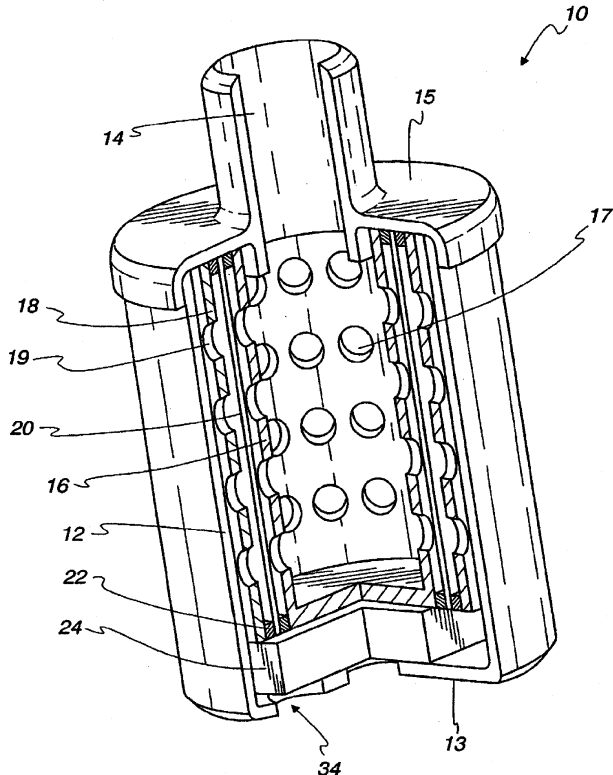
hearing aid applications. A cylindrical form allows the most effective use of available space.—JME

7,045,768

43.38.Ar DISPLACEMENT/QUANTITY OF LIGHT CONVERTER

Toru Shinzou and Yoshio Sakamoto, assignors to Kabushiki Kaisha Kenwood
16 May 2006 (Class 250/229); filed in Japan 20 February 2001

The patent describes a geometric arrangement whereby light **2** impinging on a curved surface can be reflected selectively to a receiver **3**, allowing
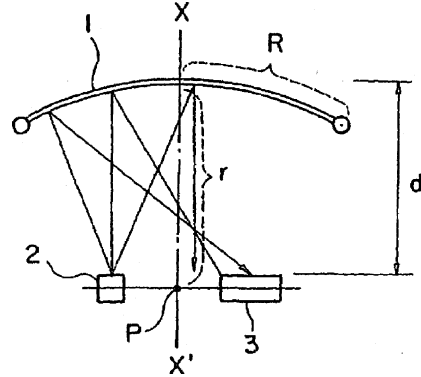
the conversion of "small mechanical displacement into a significant variation of light quantity." The idea has much in common with an optical microphone developed by Sennheiser Electronics some years ago.—JME

7,048,699

43.38.Ar NON-CYLINDRICAL ACOUSTIC WAVE DEVICE

Moshe Ein-Gal, Ramat Hasharon, Israel
23 May 2006 (Class 601/2); filed 12 September 2001

Shapes of lithotripter **10**, transducer **12**, and reflector **16** are claimed. The preferred shape for transducer **12** is noncylindrical, but conical or curved conical sections. The preferred reflector shape is also not cylindrical,
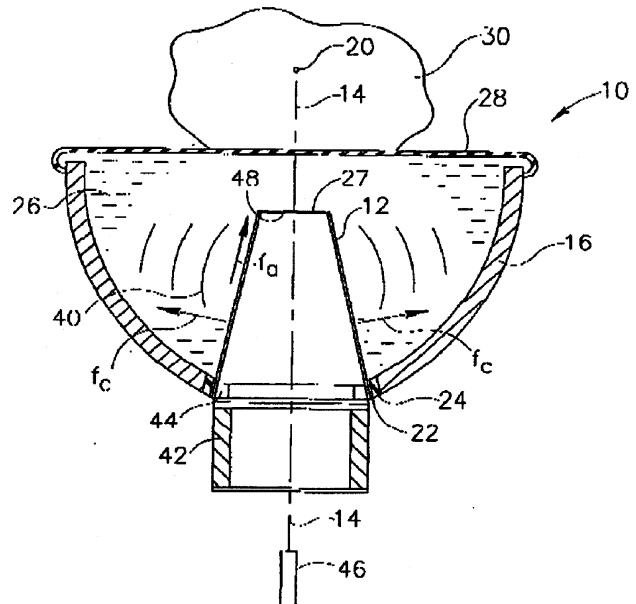
but curved conic sections. The transducer types claimed are impinging mass, explosive, electrical discharge, piezoelectric, or pneumatic.—AJC

2398   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2398

7,049,732

43.38.Ar ELECTROACTIVE POLYMERS

Qibing Pei *et al.*, assignors to SRI International
23 May 2006 (Class 310/800); filed 16 July 2004

This patent describes many different configurations of an "electroactive polymer" transducer having a built-in prestress. The configurations shown are sometimes novel, sometimes pedestrian. The discussion is very superficial; it glosses over the relative merits of the different configurations, merely cataloging individual applications. It seems that most of these were conceived previously elsewhere.—JAH

7,050,597

43.38.Ar DIRECTIONAL CAPACITOR MICROPHONE

Hiroshi Akino, assignor to Kabushiki Kaisha Audio-Technica
23 May 2006 (Class 381/356); filed in Japan 19 September 2003

The patent relates to manufacturing methods in miniature directional capacitor microphones, specifically in the area of controlling the acoustical
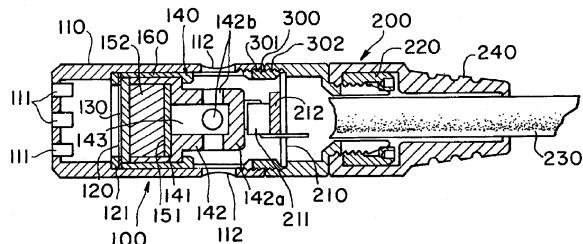
impedance of the back opening in the microphone to ensure the integrity of pattern control.—JME

7,053,523

43.38.Bs LATERAL FIELD EXCITATION OF BULK ACOUSTIC WAVES FROM AN IC-COMPLIANT LOW VOLTAGE SOURCE

Arthur Ballato *et al.*, assignors to The United States of America as represented by the Secretary of the Army
30 May 2006 (Class 310/313 B); filed 2 February 2004

This patent discloses the use of dielectric coatings over interdigital transducers (IDTs) to reduce the likelihood of electrical breakdown at the edges of the IDT fingers. There are a few other tricks that are disclosed in the patent, such as mentioned here, but this all seems like well-trodden turf.—JAH

7,053,531

43.38.Bs COMPOSITE PIEZOELECTRIC VIBRATOR

Makoto Chisaka and Osamu Kobayashi, assignors to Tayca Corporation
30 May 2006 (Class 310/334); filed in Japan 26 April 2002

The patent discloses a relatively straightforward approach to making a 3-1 composite piezoceramic from a slab of piezoceramic and epoxy filled with plastic micro-balloons. If all of this sounds familiar, it is. Only the materials have been changed.—JAH

7,024,014

43.38.Dv MULTIPLE VOICE-COIL CONE-DRIVER

Michael Noll, assignor to Harman International Industries, Incorporated
4 April 2006 (Class 381/401); filed 4 June 2003

By using separate voice coils wound on a single voice-coil former, an increase in the high-end response of an electrodynamic loudspeaker is achieved. The patent states that a third voice coil connected in parallel to the first and second voice coils, also connected in parallel, can increase the high-end response by 500 to 1000 Hz. Adding a fourth voice coil can yield
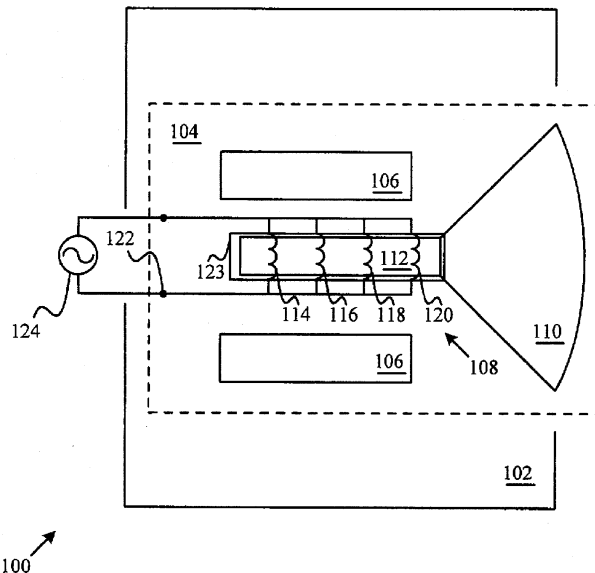
a similar increase in performance. The voice coils can be wound using bifilar, trifilar, and quadfilar techniques. The impedance of the coils is varied so that, as coils are added, the overall impedance remains the same. Some manufacturers may wish to examine claim 5, of 20, which deals with the simultaneous winding of at least two voice coils.—NAS

7,042,312

43.38.Hz LADDER-TYPE BULK ACOUSTIC WAVE FILTER WITH COMMON GROUND INDUCTOR

Sang Chul Sul and Jea Shik Shin, assignors to Samsung Electro-Mechanics Company, Limited
9 May 2006 (Class 333/187); filed in the Republic of Korea 7 October 2003

This patent describes a ladder-type filter constructed of multiple resonators in series and in parallel having a common-ground inductor. The filters are mostly band pass in the 1.8–2.0-GHz range. There is nothing novel in this class of filters.—JAH

7,046,812

43.38.Hz ACOUSTIC BEAM FORMING WITH ROBUST SIGNAL ESTIMATION

Gregory P. Kochanski and Man M. Sondhi, assignors to Lucent Technologies Incorporated
16 May 2006 (Class 381/92); filed 23 May 2000

This relatively concise patent deals with adaptive beam forming. As stated in the abstract: "Audio signals from any array of microphones are individually filtered, delayed, and scaled in order to form an acoustic beam
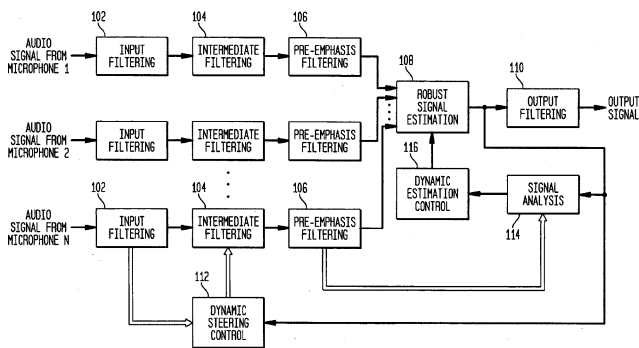
that focuses the array on a particular region. Nonlinear robust signal estimation processing is applied to the resulting set of audio signals to generate an output signal for the array. ... The nonlinear robust signal estimation processing effectively discriminates against noise originating at an unknown location outside of the focal region of the acoustic beam."—JME

7,046,816

43.38.Ja COINCIDENT SOURCE STEREO SPEAKER

Richard J. Vandersteen, Hanford, California
16 May 2006 (Class 381/182); filed 31 January 2002

The explanatory text of this patent includes numerous misleading statements and false assertions. Since the inventor is a respected loudspeaker designer, these may have originated as part of the effort to obtain a patent. Fortunately, the invention is easy to describe. Most home theater systems do not use a perforated projection screen, thus requiring the center speaker to be located above or below the screen. To achieve the desired wide, thin
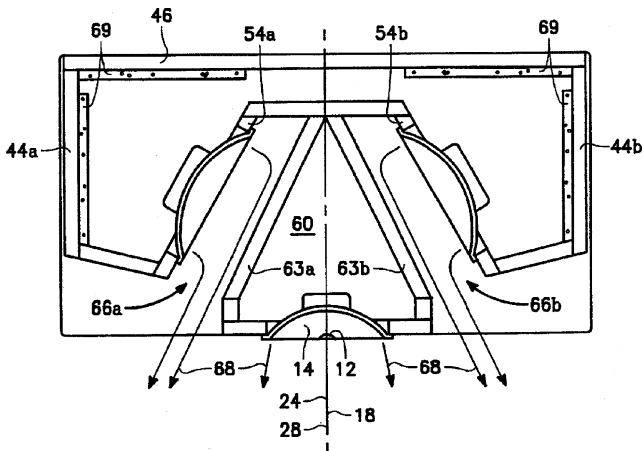
proportions, speaker manufacturers often resort to a horizontal array, consisting of a high-frequency transducer flanked by a pair of small woofers. Such an array can be designed to exhibit the same on-axis response as the left and right speakers, but off-axis response suffers. By placing woofers **66a** and **66b** as shown, their effective spacing can be reduced by half or more, at the expense of new waveguide resonances. Skeptics may argue that this arrangement simply trades one irksome problem for another.—GLA

7,050,601

43.38.Ja VOICE COIL OF SPEAKER

Takashi Suzuki et al., assignors to Matsushita Electric Industrial Company, Limited
23 May 2006 (Class 381/409); filed in Japan 19 July 2002

A loudspeaker voice coil **70** is usually wound of solid copper or aluminum wire that is connected to separate flexible leads **12** joined to fixed terminals **90**. The flexible leads may consist of several strands of fine wire
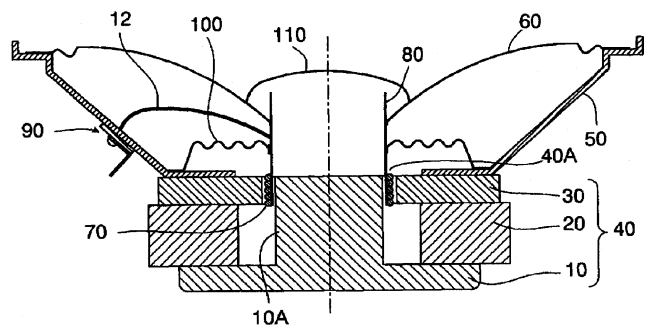
surrounding a core thread made of cotton or other fiber. This patent suggests that a simpler assembly can be realized if the voice coil itself is wound of such flexible stranded wire. "Thus, reliability is ensured and a superb sound quality is achieved."—GLA

7,016,509

43.38.Lc SYSTEM AND METHOD FOR VARYING LOW AUDIO FREQUENCIES INVERSELY WITH AUDIO SIGNAL LEVEL

Sunil Bharitkar and Timothy Shuttleworth, assignors to Harman International Industries, Incorporated
21 March 2006 (Class 381/98); filed 8 September 2000

A means is disclosed of adjusting a loudness-compensation filter based on the signal level and the inverse of the Fletcher-Munson curves. The technique can be implemented in both the analog and digital domains. The patent is well written and easy to follow, although better proofreading would have caught the problems with the labeling of the curves in Fig. 2 that renders it almost useless, as well as some nomenclature discrepancies in the text.—NAS

7,038,534

43.38.Lc CLASS D AUDIO AMPLIFIER AND METHOD FOR COMPENSATION OF POWER SUPPLY VOLTAGE INFLUENCE ON OUTPUT AUDIO SIGNAL IN CLASS D AUDIO AMPLIFIER

Tomasz Hanzlik, assignor to Advanced Digital Broadcast Polska Sp. z o.o.
2 May 2006 (Class 330/10); filed in Poland 21 January 2002

At the present time, there is considerable interest in developing high-quality class D audio amplifiers because very high conversion efficiency is possible. However, the interface between the amplifier output and the loudspeaker is finicky, and it is difficult to utilize conventional negative feedback
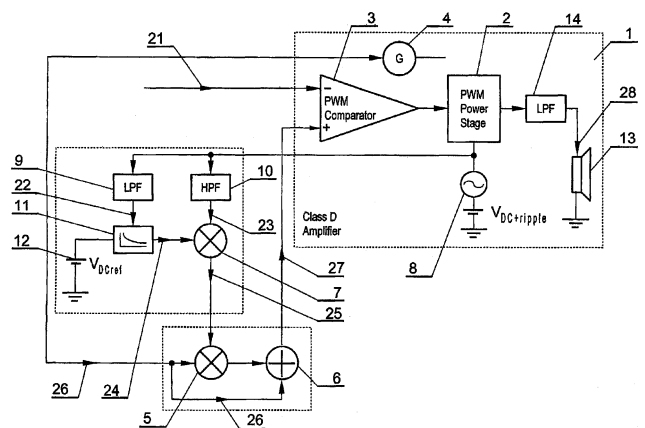
for distortion reduction. Yet without some form of negative feedback, any ripple in the power supply modulates the signal. Alternative schemes have been patented that attempt to correct the PWM signal in reaction to power supply voltage fluctuations. This short patent discloses such a scheme. Generator **4** produces sawtooth or triangular waves. The power-supply voltage, including ripple **8**, is fed to separate filters, lowpass **9** and highpass **10**, deriving a rapidly fluctuating signal and a slowly fluctuating signal. These are combined with the generator output through multiplier **5** and adder **6** to create a predistorted carrier signal **27**. The patent is well written and the invention is clearly disclosed in both the specifications and the claims sections.—GLA

7,046,814

43.38.Lc IMPEDANCE ADAPTER PREAMPLIFIER DEVICE FOR ELECTRONIC TUBE AUDIO FREQUENCIES ABLE TO BE INSERTED ON LINE ON THE PATH OF THE LOW FREQUENCY SIGNAL

Nicolas Girard, Paris, France
16 May 2006 (Class 381/120); filed in France 15 September 1999

A simple vacuum-tube line amplifier might consist of a cathode follower driven by a triode gain stage, providing very high input impedance and low output impedance. Variations of this well-known configuration have been in use for more than 50 years, including portable devices utilizing
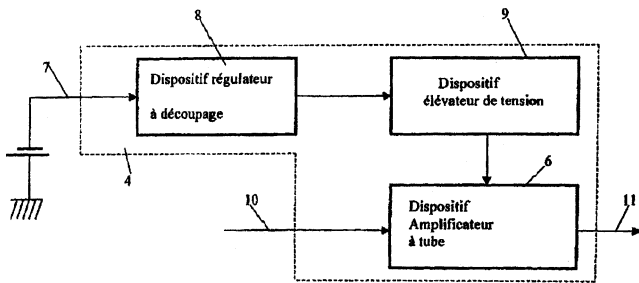
small "A" and "B" batteries. If a switching-type power supply were used, the "B" battery could be eliminated, and that appears to be the sole novel feature of this invention.—GLA

7,038,575

43.38.Md SOUND GENERATING APPARATUS FOR USE WITH GLOVES AND SIMILAR ARTICLES

Jan A. Frohman and James L. Jorgensen, assignors to The Board of Regents of the University of Nebraska
2 May 2006 (Class 340/384.7); filed 31 May 2001

Simply, the inventors propose adding the ubiquitous sound chip to a pair of gloves for use at a (cold) sporting event. Mechanical actuators (i.e., switches) will turn on the "sound generation apparatus" (speaker **24**). What
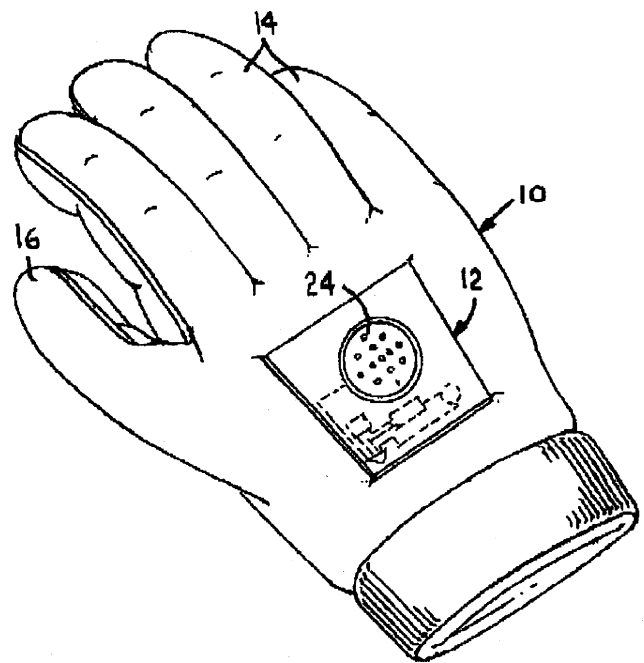
seems to have been forgotten is the audio power required in a crowd and the size of the accompanying battery.—MK

7,038,606

43.38.Md DIGITAL SIGNAL PROCESSING DEVICE AND AUDIO SIGNAL REPRODUCTION DEVICE

Tsunehiko Hongoh and Hirotoshi Yamamoto, assignors to Sharp Kabushiki Kaisha
2 May 2006 (Class 341/143); filed in Japan 29 October 2002

Like the DASD scheme, the inventors propose using an oversampled one-bit stream with a seventh-order delta-sigma modulator **5**. The output
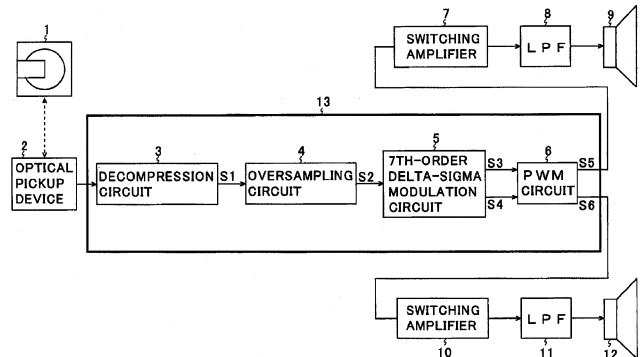
of the modulator is amplified by a PWM (class D) amp of unspecified design. After (passive) lowpass filtering, the output is suitable for loudspeakers.—MK

7,043,027

43.38.Md SYSTEM FOR DETECTING FAILURES IN A LOUDSPEAKER ASSEMBLY

Bradley C. Wood, assignor to Harman International Industries, Incorporated
9 May 2006 (Class 381/58); filed 1 October 2001

It is usual for surround sound decoders and computer sound cards to provide a diagnostic check in which a test signal is sent sequentially to all channels—left, center, right, left surround, etc. But suppose that the sound
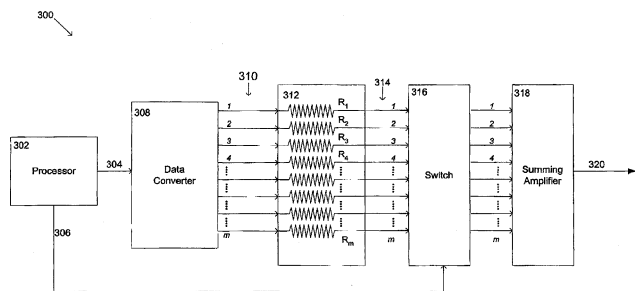
card itself is defective? This patent describes a similar but very specific method for sending a test signal sequentially to all loudspeakers in a multimedia system. Since the associated circuitry is to be included as part of the speaker kit, any problems can be isolated from other components.—GLA

7,030,539

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE AND ELECTRONIC COMPONENT COMPRISING IT

Mitsuhiro Furukawa *et al.*, assignors to Matsushita Electric Industrial Company, Limited
18 April 2006 (Class 310/313 D); filed in Japan 3 October 2001

Metal film can be cut to SAW pattern **2**, **3** by an ion beam that also leaves a charge that bends said beam while cutting the deeper metal. Carbonized photoresist is deposited in the cut region that reduces the acoustic
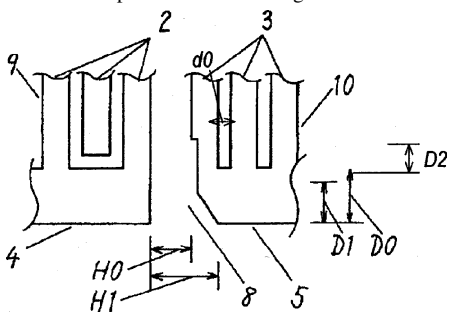
energy produced in use. To prevent this anomaly, a portion **5** of the reflector electrode set **3** is cut off.—AJC

7,034,433

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Michio Kadota *et al.*, assignors to Murata Manufacturing Company, Limited
25 April 2006 (Class 310/313 A); filed in Japan 12 October 2001

The author claims a surface acoustic wave device **11** that has high electromechanical and reflection coefficients. The chosen material **12** is
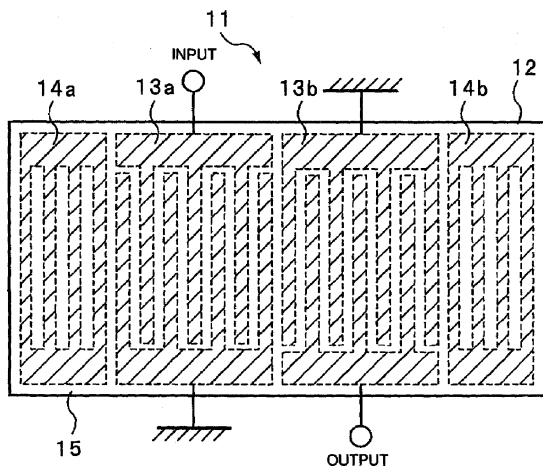
25°–50° Y-plate lithium tantalate with preferred Euler angles, having aluminum interdigital transducer (IDT) fingers **13** and **14**, with optimized aluminum thickness and with silicon oxide film **15** covering at least one IDT.—AJC

7,034,434

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Koji Yamamoto *et al.*, assignors to Murata Manufacturing Company, Limited
25 April 2006 (Class 310/313 R); filed in Japan 3 June 2002

SAW chip **2** and substrate **3** are held separate by gold bump **4** so that vibration space **7** is assured. Resin seal material **5** is prevented from flowing
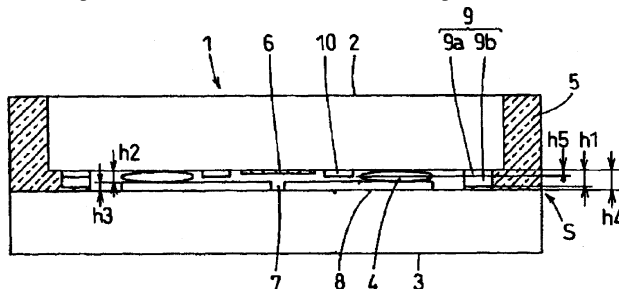
into space **7** by barrier bump **9**. The authors claim various groove and bump geometries.—AJC

7,038,353

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE AND METHOD OF MANUFACTURING THE SAME

Yoshio Maeda *et al.*, assignors to Seiko Epson Corporation
2 May 2006 (Class 310/313 D); filed in Japan 10 January 2001

For low-frequency (viz. 100 MHz) SAW operation, Rayleigh wave reflection must occur in order that the device be efficient and also small. Rayleigh waves created between IDT electrodes **14a** and **14b** flow to the
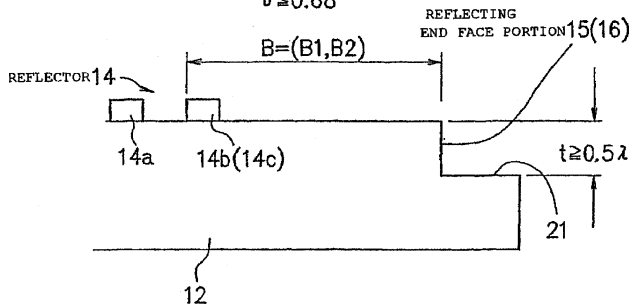
right and are reflected by end face **15** placed at the favored distance **B**, where δ is an allowance determined from the number of IDT reflector strips.—AJC

7,038,560

43.38.Rh SURFACE ACOUSTIC WAVE FILTER AND SURFACE ACOUSTIC WAVE RESONATOR HAVING BENT ELECTRODE FINGERS

Katsuro Yoneya, assignor to Seiko Epson Corporation
2 May 2006 (Class 333/195); filed in Japan 19 June 2003

A surface acoustic wave filter with bent fingers is claimed that suppresses spurious higher frequency modes, or moves their frequency outside of the useful pass band of the filter. Fingers **121**, **122**, **131**, **132**, **141**, and **142**
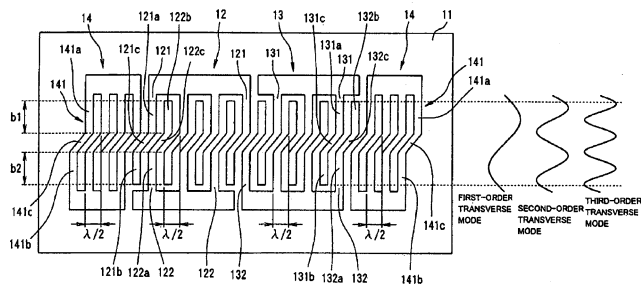
are shaped into three or more segments **a**, **b**, and **c**. Segment **c** is as short as possible and arranged to produce a specific side shift **L**, shown here, for instance, as $\frac{1}{2}$ surface wave wavelength. "[R]esonance in the first-order transverse mode...becomes primary response (sic) ... ." Resonant energy in the second-order mode cancels, and resonant energy in the third-order mode becomes the main spurious response, but at a frequency sufficiently changed as to be outside the normal pass band of the filter. The claims include multiple segments (zig-zag) and different shifting distance **L**.—AJC

7,042,056

43.38.Rh CHIP-SIZE PACKAGE PIEZOELECTRIC COMPONENT

Yoshihiro Koshido, assignor to Murata Manufacturing Company, Limited
9 May 2006 (Class 257/414); filed in Japan 31 July 2002

Surface acoustic wave (SAW) resonator **204-205** is protected from deterioration in use by providing a closed operating space between bonding substrate **223** and piezoelectric substrate **1**. External terminals **226** and **227** are provided by solder **226a** and **227b** through bonding substrate **223a** to
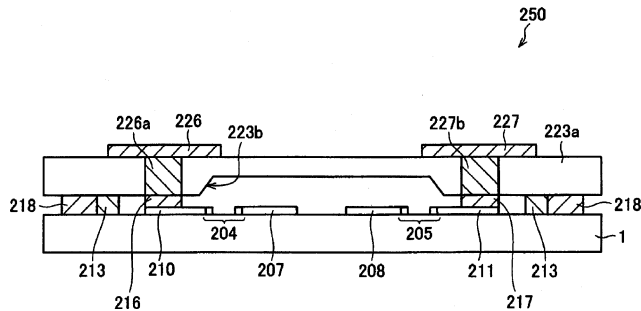
wiring lines **210** and **211**. A recess **223b** assures a clear and corrosion-free space over the SAW IDT fingers **207** and **208**. The claims include fabrication sequences and solder alloys. The fifth of eight embodiments is shown.—AJC

7,042,131

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Yuichi Takamine, assignor to Murata Manufacturing Company, Limited
9 May 2006 (Class 310/313 R); filed in Japan 6 June 2001

A 36° Y-cut, X-propagating lithium tantalate substrate **2** has tandem pass-band SAW filter elements **3** and **4**. Silicon dioxide film **7** is deposited over at least one SAW terminal pair, with film thickness optimized for the
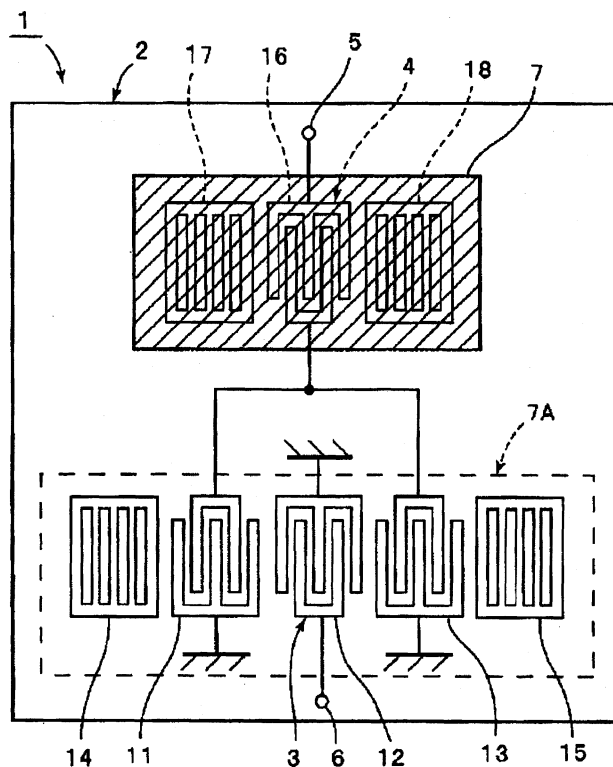
(aluminum) finger thickness and to present a minimal temperature coefficient of the frequency of the band-pass edges.—AJC

7,042,133

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE AND METHOD OF ADJUSTING A TEMPERATURE CHARACTERISTIC OF THE SAME

Shigeo Kanna, assignor to Seiko Epson Corporation
9 May 2006 (Class 310/313 A); filed in Japan 4 October 2002

Two surface acoustic wave devices **M1** and **M2** on quartz piezoelectric substrate **3**, cut at specific Euler angles, propagate Rayleigh surface waves in two slightly different directions $\psi 1$ and $\psi 2$, whose wave velocities have
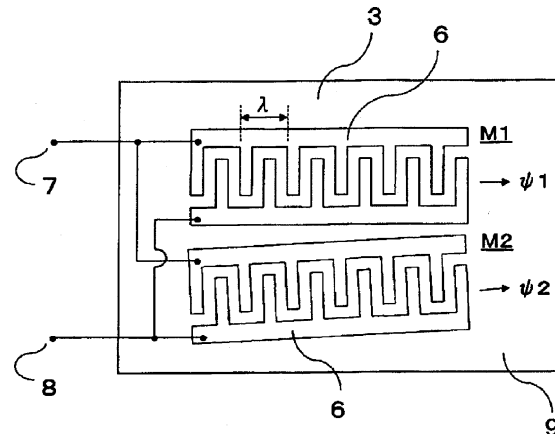
temperature coefficients that differ slightly. A calculation method is provided to select the optimum propagation directions $\psi 1$ and $\psi 2$, especially selecting the turnover temperatures with respect to the operating temperature range.—AJC

7,046,102

43.38.Rh SURFACE ACOUSTIC WAVE FILTER, BALANCED TYPE CIRCUIT, AND COMMUNICATION APPARATUS

Hiroyuki Nakamura *et al.*, assignors to Matsushita Electric Industrial Company, Limited
16 May 2006 (Class 333/133); filed in Japan 6 March 2002

The uniformity (flatness) of the passband of a rf surface acoustic wave filter **A-A′** is said to be improved by adding a thin layer of lower dielectric-constant film **201′** (e.g., silicon nitride having a dielectric constant of 10)
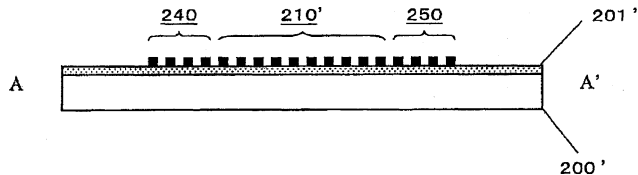
over the piezoelectric substrate **200′** (e.g., lithium tantalate of dielectric constant 40) before the fingers **240-210′-250** are applied.—AJC

7,039,179

43.38.Si ECHO REDUCTION FOR A HEADSET OR HANDSET

John F. Gerhardt *et al.*, assignors to Plantronics, Incorporated
2 May 2006 (Class 379/390.02); filed 27 September 2002

There must be more than 100 U.S. patents dealing with ways to reduce echo in telephone communications, yet this patent cites only one reference and that was supplied by the examiner. Although this reviewer is not an expert in the field, the patent claims suggest that the core of the invention is
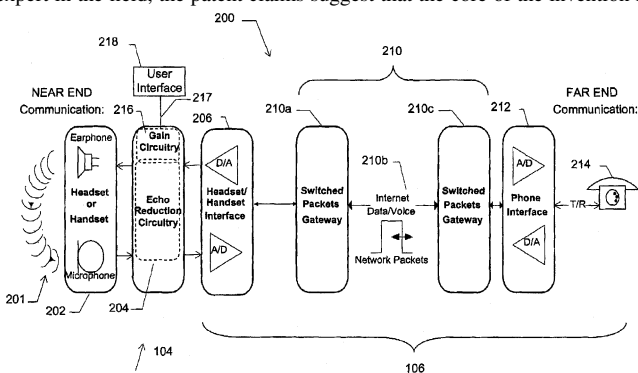
not a new echo cancelling circuit, but rather a practical method of achieving useful echo cancellation when communicating over packet-switched networks. A variety of handsets and headsets can be accomodated, allowing the user to enjoy improved echo reduction without replacing every piece of communications equipment.—GLA

7,043,279

43.38.Si HANDSFREE STRUCTURE WITH ANTIBACKGROUND NOISE FUNCTION

Ching-Chuan Lee, Taipei, Taiwan
9 May 2006 (Class 455/569.1); filed in China 3 December 2002

A variety of small, plug-in headsets are available to cellular phone users, allowing hands-free operation. The headset patented here adds a

separate vibration pickup (throat microphone) that can be used in a high-noise environment.—GLA

7,043,286

43.38.Si EXTRINSIC SIGNAL TO SHUNT AN ACOUSTIC DRIVER IN A CELLULAR TELEPHONE, PAGER OR THE LIKE

David Leason, assignor to Leason Holding Company, LLC
9 May 2006 (Class 455/701); filed 13 June 2001

Wouldn't it be nice if cellular phone ring signals could be automatically disabled inside a concert hall or movie theater? This patent describes a method of achieving this goal, assuming that all cellular telephones include

the required circuitry. Since cell phones are known to be used by terrorist groups, perhaps a mandate could be enforced by the Department of Homeland Security.—GLA

7,046,082

43.38.Si ELECTRONIC DEVICE HAVING SOUND OUTPUT MODULE

Kunihiro Komiya and Tadayuki Sakamoto, assignors to Rohm Company, Limited
16 May 2006 (Class 330/51); filed in Japan 27 December 2002

The title of this fairly long patent is misleading; it actually describes transient suppression circuitry for use in audio amplifiers. When a quiescent output circuit is energized by switching on its bias voltage, an audible pop results. If a simple R-C filter is used to control charging time, it may not provide sufficient power-supply rejection. The invention adds a fairly complicated "bias adjustment circuit" to provide optimum voltage rise characteristics.—GLA

2404   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2404

7,047,052

43.38.Si CELLULAR PHONE TERMINAL

Hiroshi Akai *et al.*, assignors to Hitachi, Limited
16 May 2006 (Class 455/575.1); filed in Japan 19 July 2002

Laptop DVD players include loudspeakers for stereo sound, so why can't we enjoy stereo music downloads on our cell phones? In fact, suitable technology has already been patented. Well, if a cell phone has a video

screen and stereo speakers, then the next logical step is video teleconferencing with stereophonic audio pickup and reproduction. That is the challenge taken up by this patent. A cellular phone is described that houses a "plurality of speakers" and a "plurality of microphones." It may also include a proximity sensor to automatically switch between normal and stereo pickup.—GLA

7,050,796

43.38.Si METHOD OF CONTROLLING THE AUDITORY RESPONSE OF WIRELESS DEVICES

Thomas Ward Humphrey and Larry Dean Moore, assignors to
International Business Machines Corporation
23 May 2006 (Class 455/420); filed 25 January 2002

IBM seems to be interested in what might be called semi-intelligent local networks. With such a network, if your daily schedule is entered into your desktop computer, it automatically becomes available through your

palmtop computer, cell phone, etc. Moreover, the behavior of individual devices can be modified by their location or, as this patent suggests, by the ambient sound field. For example, if the environment is quiet and other wireless devices are present, the network may decide that you are in conference and disable your cell phone except for emergency messages.—GLA
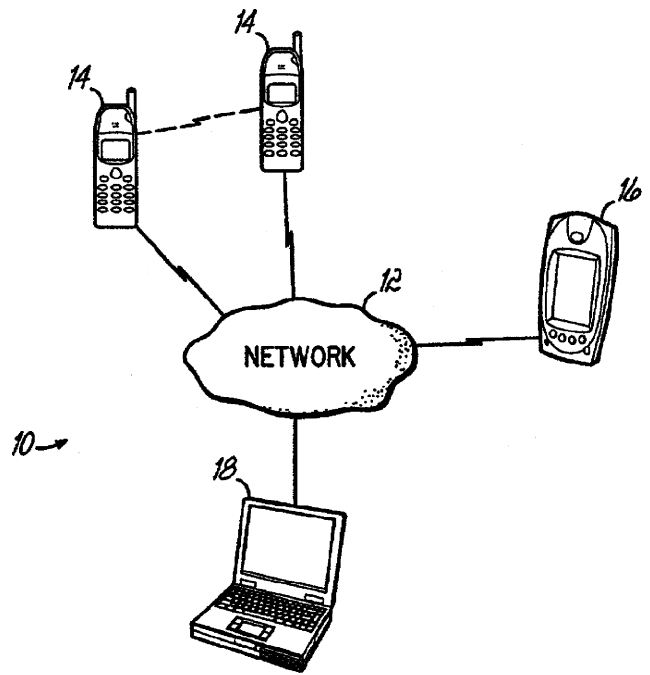
7,013,012

43.38.Tj METHOD FOR PROCESSING SOUND EFFECTS IN A NOTEBOOK COMPUTER

Li-Chun Lo, assignor to Quanta Computer Incorporated
14 March 2006 (Class 381/307); filed in Taiwan
20 September 2002

This patent describes a scenario wherein a determination is made as to whether an audio signal is a stereo signal or a 4/5.1-channel signal. For stereo signals, the internal speakers of the notebook computer are used. For four-channel signals, the left and right signals are presented for use via the
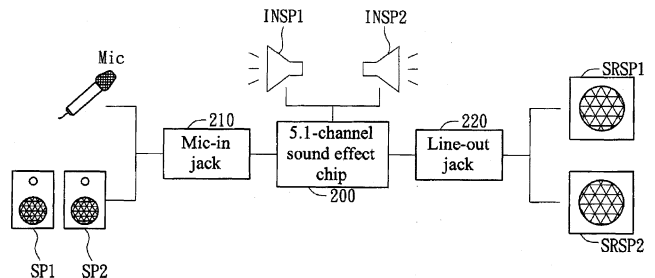
mic-in jack and the surround channels via the line-out jack. For 5.1 signals, the mic-in jack is used for the subwoofer (or woof as it is called in the patent) and the center channel with the line-in jack signals as before. Sounds like a lot of jack switching and outboard equipment may be needed in addition to the "laptop" computer.—NAS

43.38.Tj PORTABLE INFORMATION PROCESSING APPARATUS

Dai Ueda *et al.*, assignors to Matsushita Electric Industrial Company, Limited
21 March 2006 (Class 361/683); filed in Japan 28 March 2003

A means of mounting piezoelectric audio transducers **5** in the top cover **9** of a notebook computer **1** is described. The goal is to save space and protect the speakers from rain or dust. Protection of the keyboard and
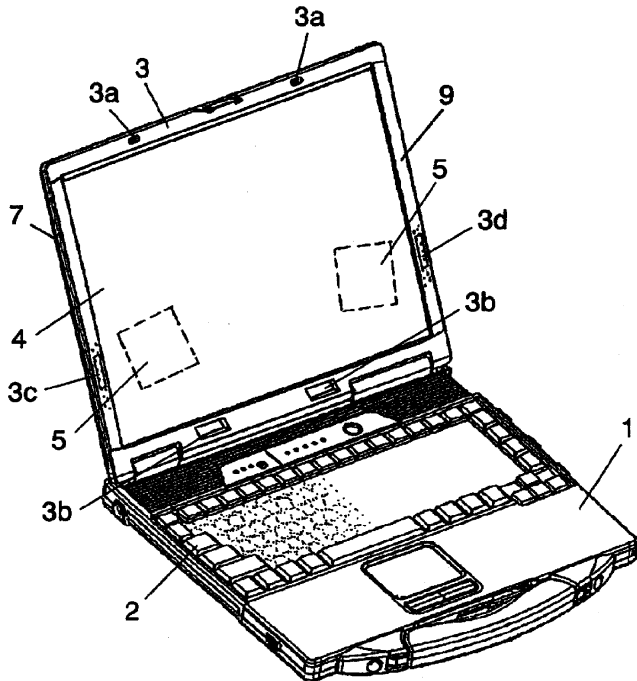
display from these contaminants is not discussed, although the invention may be destined for use in the assignee's line of Panasonic "ruggedized" notebook computers.—NAS

43.38.Tj SHARING DEVICE FOR MULTIPLE AUDIO SOURCES

Ying-Cou Wang, Taipei, Taiwan
21 March 2006 (Class 369/2); filed 26 November 2002

Many new homes (and older ones so fitted) now have in-house audio and video systems that take multiple sources and allow for their distribution to multiple outputs (rooms). The patent describes a way of doing this that
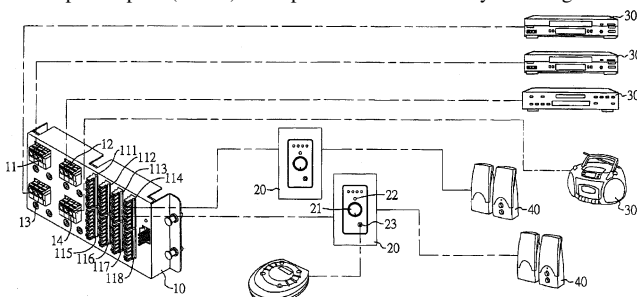
many may recognize as being a good description of the prior art perhaps installed in their own home or easily available as an option in homes now offered for sale by some of this nation's home-building conglomerates.—NAS

43.38.Tj VARIABLE DISPOSITION WIDE BAND MULTI-WAY LOUDSPEAKERS

Guido Noselli, assignor to Outline S.n.c. di Noselli G. & C.
21 March 2006 (Class 381/386); filed in Italy 2 March 2000

The main point of the patent is that the internal volume of subwoofer and woofer loudspeaker systems can be used to nest the smaller components of a large system while transporting same, so as to use what would be wasted space. For example, the mid-hi enclosure **11** can fit inside the subwoofer enclosure **10** when traveling. Various combinations and permutations
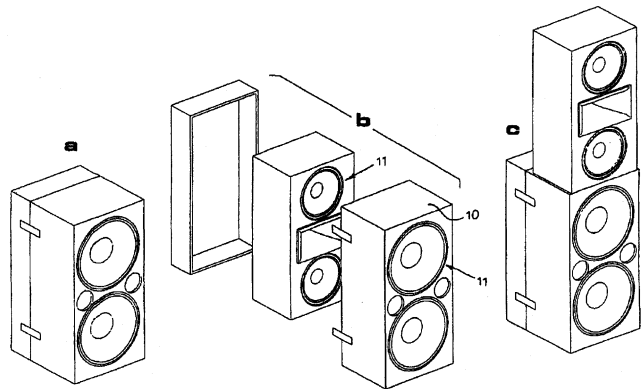
of the nesting (nestee?) box and nestor box, as well as means for actuating the deployment of the inner box, are described and illustrated. Questions about resealing the bigger box when the rear cover is replaced, as well as how stiff the enclosure can be without interfering with the nesting capability, are not discussed.—NAS

43.38.Vk ACOUSTIC CORRECTION APPARATUS

Arnold I. Klayman and Alan D. Kraemer, assignors to SRS Labs, Incorporated
9 May 2006 (Class 381/98); filed 22 January 2004

The patent assignee has long been at the forefront of two-channel matrix enhancement techniques for a variety of consumer applications,
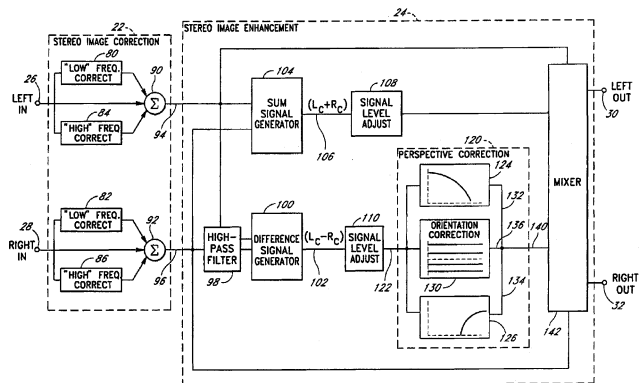
principally television sound transmission and playback in the home. The patent covers ongoing improvements in image definition and stability.—JME

7,043,036

43.38.Vk AUDIO REPRODUCING APPARATUS

Yuji Yamada and Koyuru Okimoto, assignors to Sony Corporation
9 May 2006 (Class 381/309); filed in Japan 18 March 2002

This patent describes a number of signal processing techniques for converting one or more audio signals for playback over headphones. The problems of "in-the-head" localization are addressed through multiple
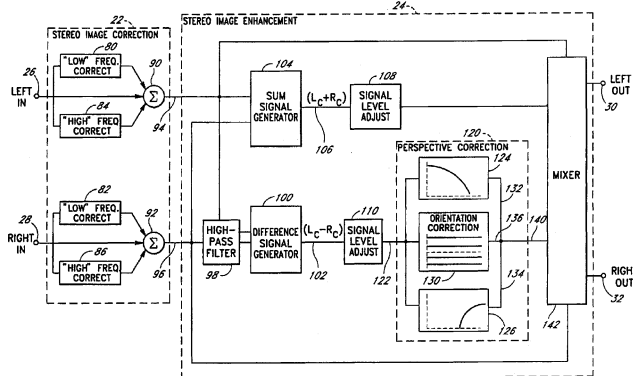
delays and proper equalization, presumably resulting in multiple images that can be positioned around the listener. The figure shows how a single input channel may be processed for this purpose.—JME

7,050,849

43.40.Ng VIBRATIONAL THERAPY DEVICE USED FOR RESYNCHRONIZATION PACING IN A TREATMENT FOR HEART FAILURE

Debra S. Echt et al., assignors to EBR Systems, Incorporated
23 May 2006 (Class 607/3); filed 15 June 2004

This system for pacing the heart includes a vibrational transducer that directs energy to the heart, usually aimed at a ventricle, to pace the heart and to promote synchronized contraction of the ventricles. Also, additional
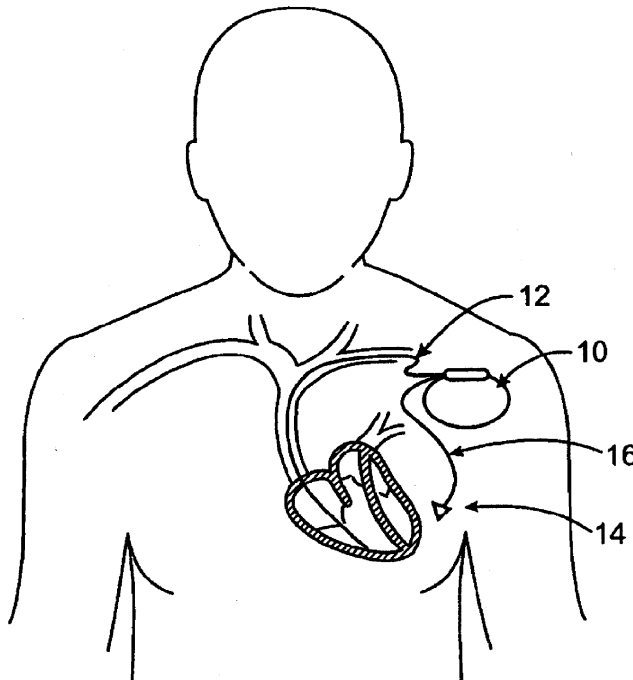
vibratory and/or electrical stimulation may be provided. These vibrational transducers are usually implanted near the vicinity of the heart.—DRR

7,038,357

43.40.Sk STRETCHED ROLLED ELECTROACTIVE POLYMER TRANSDUCERS AND METHOD OF PRODUCING SAME

Andrew A. Goldenberg et al., assignors to Engineering Services Incorporated
2 May 2006 (Class 310/328); filed 21 August 2003

This actuator transducer is a cylindrical tube (axis **114**) comprising stretched rolled electroactive polymers. The wall **102** is laminated **132–138**. Thin electrostriction film **132** and conducting film **134** are wrapped in the
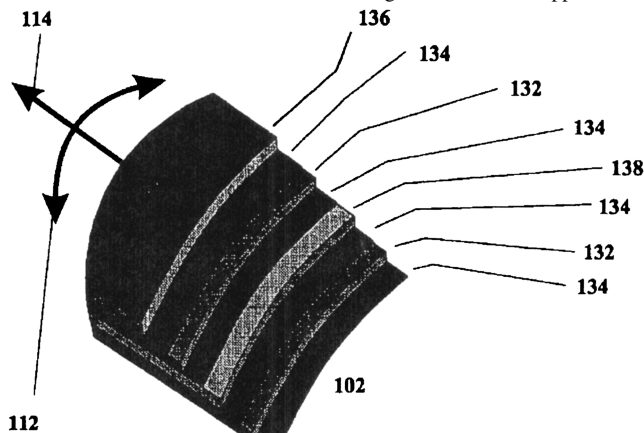
circumferential direction on substrate **138**. The tube interior can house electronics within and mechanical connections and terminals on its ends. Variations include patterned regions of electrostriction and strain feedback. Unfortunately, specific forces and possible actions are not detailed in the text.—AJC

7,048,473

43.40.Tm VIBRATION-PROOF CONSTRUCTION METHOD

Hirokazu Takemiya, assignor to Hirokazu Takemiya
23 May 2006 (Class 405/302.5); filed in Japan 5 November 2002

This patent describes means for attenuating vibration transmission through the ground—for example, from railways to nearby buildings. The idea in essence involves the introduction of alternate stiff and resilient layers along a portion of the propagation path. A preferred embodiment makes use of arrays of closely spaced or contiguous concrete piles surrounded by rubber from pulverized scrap tires. Other embodiments employ concrete plates with rubber along their surfaces.—EEU

7,025,342

43.40.Vn ACTUATOR DRIVE CONTROL DEVICE FOR ACTIVE VIBRATION ISOLATION SUPPORT SYSTEM, AND METHOD OF USING SAME

Hirotomi Nemoto et al., assignors to Honda Motor Company, Limited
11 April 2006 (Class 267/140.14); filed in Japan 13 June 2003

This is related to United States Patent 7,010,889 by Atsusi Abe. Here, the angular acceleration of the crankshaft is also computed and applied as a correction to the control signal to this actuator.—AJC

43.40.Vn APPARATUS AND METHOD FOR DETECTING AND COMPENSATING FOR RESONANCE FREQUENCY IN DATA STORAGE SYSTEM

Da-woon Chung, assignor to Samsung Electronics Company, Limited

18 April 2006 (Class 360/75); filed in the Republic of Korea 10 November 2003

The resonant frequency and response of a hard-disk, read-head-servo drive system is determined in two steps: First, a gain increase of the position-error-signal (PES) amplification is made in boosting filter **203** driving the voice-coil-motor (VCM) actuator **206** so that resonant oscillation
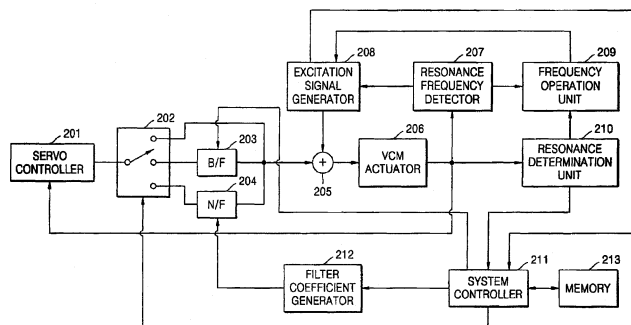
occurs. Resonant frequency detector **207** detects that frequency. Second, the VCM is fed a signal having the detected resonant frequency, and the PES is analyzed for a possible second resonant frequency. Information on these frequencies is analyzed by **210** and the results stored in **213** for subsequent use in **211** and **212**.—AJC

7,046,478

43.40.Vn METHOD AND APPARATUS FOR REDUCING VIBRATION IN A DYNAMIC SYSTEM

Tao Zhang et al., assignors to Seagate Technology LLC
16 May 2006 (Class 360/77.08); filed 11 April 2003

In order to control the displacement of a dynamic system, such as a disk drive, primary position-control signals are generated at a position-measurement sampling frequency. Position-control performance is improved by the use of additional position-control signals that occur between the primary ones and whose characteristics are determined on the basis of the primary signals.—EEU

7,047,109

43.40.Vn MACHINERY FOR IMPROVING PERFORMANCE IRREGULARITIES ARISING FROM VIBRATIONS

Katsuyuki Ogura and Hirotada Anzai, assignors to Mitutoyo Corporation
16 May 2006 (Class 700/280); filed in Japan 13 May 2003

Apparatuses like coordinate measuring machines include moving platforms that are subject to performance-degrading vibrations. In order to suppress the vibrations of such a platform that moves in a given direction, an inertial actuator is attached to the platform, with the actuator's motion controlled in response to appropriately processed signals from a vibration sensor. The actuator described in this patent consists of two similar counter-rotating unbalanced flywheels that are driven by a motor that is controlled so that the net inertia force resulting from any rotating imbalance acts at the same frequency as the offending vibrations and counteracts these.—EEU

43.50.Gf SILENCER FOR VENTILATION SYSTEM AND METHODS

Leon Robert Cuvelier et al., assignors to Donaldson Company, Incorporated
23 May 2006 (Class 415/119); filed 1 March 2004

This silencer for a ventilation system (which could be the intake of a gas turbine) functions by directing flow through two body sections spaced apart from each other to define a gas flow path in between. The first body section consists of a base, a sidewall, and an upper wall. The upper wall has a center region with a concave wall smoothly sloping downward, terminating at the sidewall. The base, sidewall, and upper wall together form a
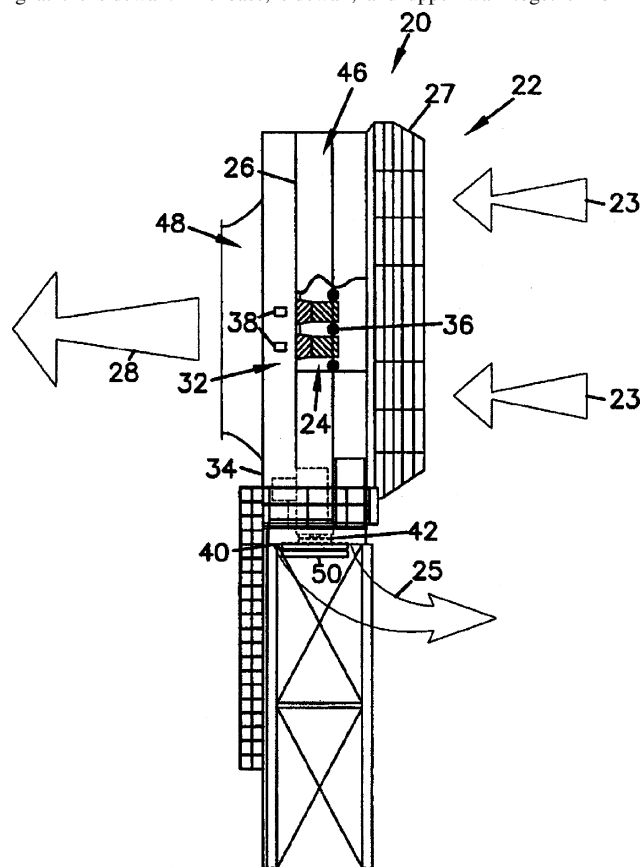
sectional interior volume that is preferably filled with a packing material. The second body section has a base, inner and outer sidewalls, and an upper wall. The base and walls together define an interior volume that is also filled with packing material. The second body section also has a center aperture, wherein the inner sidewall lines the center aperture.—DRR

7,042,218

43.50.Ki SYSTEM AND METHOD FOR REDUCING AUDITORY PERCEPTION OF NOISE ASSOCIATED WITH A MEDICAL IMAGING PROCESS

Michael B. Sellers, assignor to General Electric Company
9 May 2006 (Class 324/309); filed 6 May 2004

A noise-cancellation method is described for reducing the perception of noise generated as a byproduct of a medical imaging process. The imaging scanner is configured to scan a subject within an imaging area that emits system noise during the course of operation. An ultrasonic emitter system is incorporated to emit an inaudible signal having properties to reduce perception of the system over at least a portion of the imaging area. It is apparent

from the patent description that the noise cancellation applies to the upper frequencies, with questionable results for the audio range.—DRR

7,035,414

43.50.Lj ACTIVE NOISE ATTENUATION SYSTEM

Paul D. Daly, assignor to Siemens VDO Automotive Incorporated
25 April 2006 (Class 381/71.4); filed 28 February 2001

Vehicle-engine air intake **18** with opening **16** is fitted with a cone-shaped insert **20** faced with loudspeaker **46**. Microphone **42** on arm **54** senses emitted sound, providing a signal to electronics center **38**, which
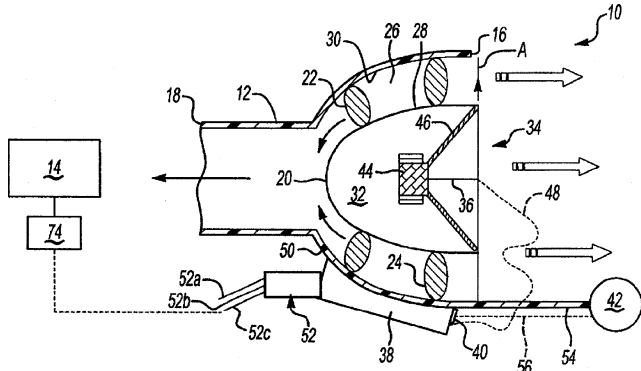
reverses its phase and feeds it back to speaker **46** to attenuate the emission of engine air-induction-system noise. Engine firing signals **52** are also fed to **38**. The author also claims the preferred forward microphone position at **42** and the use of flexible connectors **48** and **56**.—AJC

7,016,503

43.58.Vb ADAPTIVE EQUALIZER FOR VARIABLE LENGTH SOUND TUBES UTILIZING AN ACOUSTIC PRESSURE RESPONSE MEASUREMENT

Richard Lance Willis, assignor to Motorola, Incorporated
21 March 2006 (Class 381/58); filed 30 December 2002

By using the acoustic response technique (as in the figure), the electrical impedance response method, or the time of flight technique, the equalization needed to compensate for the length of tubing used in an earpiece
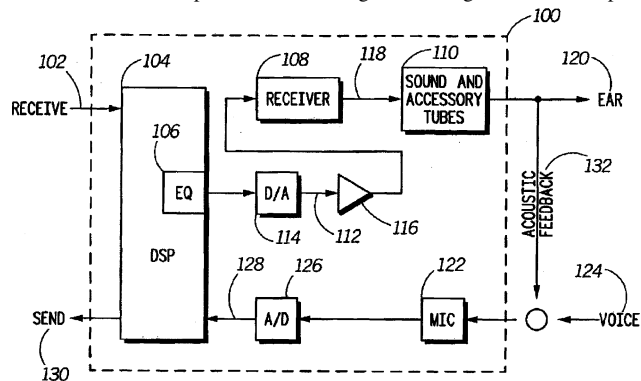
can be determined and then automatically implemented. The measurement can be used for hearing aids as well as for the plethora of blue-tooth and wired earpieces now used with cell phones and other personal entertainment devices.—NAS

7,027,921

43.60.Qv SURFACE ACOUSTIC WAVE SENSOR

Kourosh Kalantar-Zadeh and Wojtek Wlodarski, assignors to Microtechnology Centre Management Limited
11 April 2006 (Class 702/2); filed in Australia 21 May 2001

Substrate **1** supports a SAW source **3** and detector **4**. A shear wave is transmitted through a layer **5** coated with a sensing layer **6**. The sound velocity and attenuation of **5** and **6** combined will change when exposed to
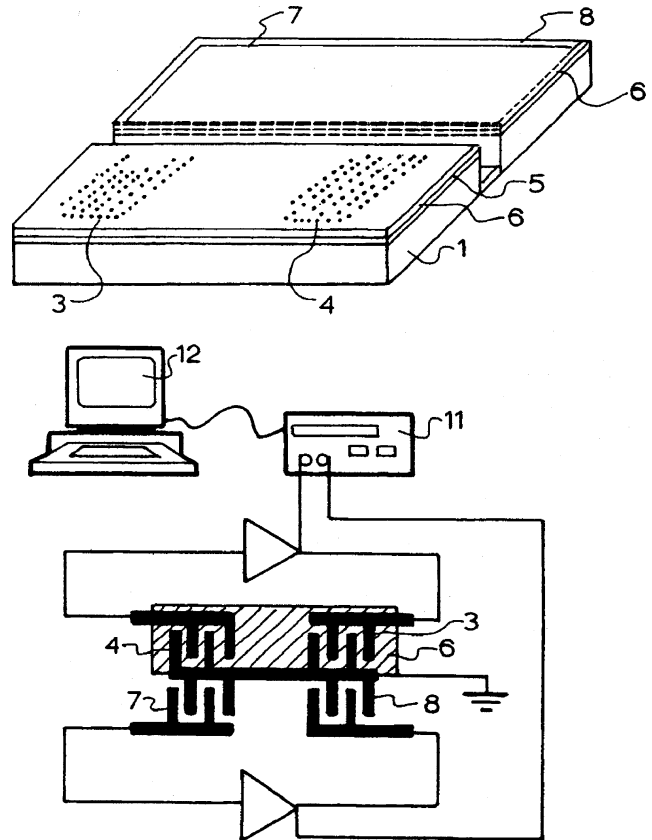
a specified gas or liquid. Frequency counter **11** and computer **12** (algorithms not specified) are used to correlate the received level with concentration.—AJC

7,028,548

43.60.Qv VIBRATION TYPE ANGULAR VELOCITY SENSOR

Hirofumi Higuchi and Takeshi Ito, assignors to Denso Corporation
18 April 2006 (Class 73/514.12); filed in Japan 31 March 2003

This device senses the vehicle turn rate (angular velocity) in the presence of random vibration and noise. Bodies **4a** and **4b** oscillate in the x direction, but with opposite phase. The vehicle turning rotates enclosure **10** to produce a coriolis force in the y direction, which in turn causes complementary signals to be generated via capacitor plates **5a** and **5b**. Linear
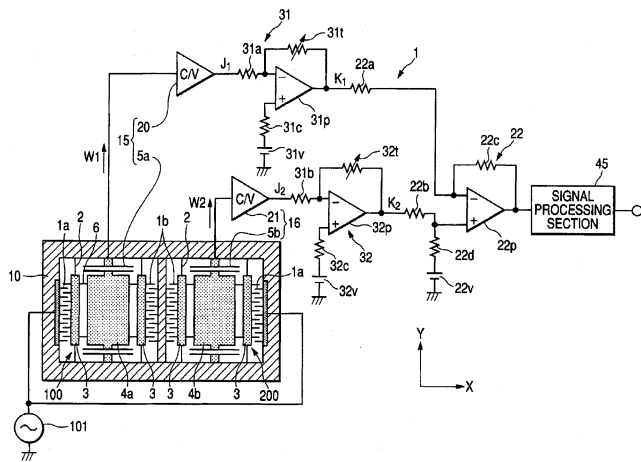
vibrations of enclosure **10** via random vehicle motions causes capacitor-plate **5a** and **5b** signals that are the same in amplitude and phase, thereby cancelling. The claims include laser trimming of gain setting resistors **31t** and **32t**.—AJC

7,038,780

43.60.Rw METHOD FOR SENSING ACOUSTIC ENERGY

Joseph Lawrence Chovan *et al.*, assignors to Lockheed Martin Corporation
2 May 2006 (Class 356/432); filed 20 June 2002

A source **S1** emits sound energy **31** into dense medium **10**, such as water. Part of that sound energy **33** continues into less dense medium **20**. The radius of curvature of wave front **33** is increased so as to be essentially parallel to the interface **25**. A laser beam source and detector **100-100** transmits a beam **110** through **20** parallel to the interface **25**. Subsequent pressure modulation by **33** of medium **20** produces a corresponding variation in
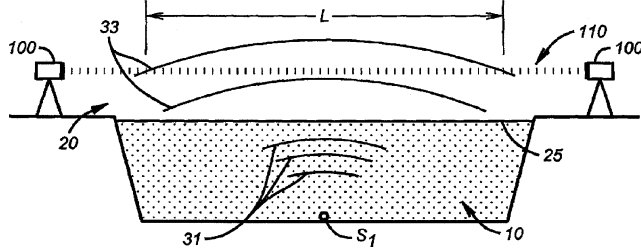
phase of the transmitted laser beam **110**, causing a phase-shift modulation of the laser energy received by detector **100**. Intermediate frequency (IF) modulation of the laser beam source serves to distinguish laser light from other light received by **100**. Sound modulation of **S1** is presented as a corresponding phase-shift modulation found by laser beam receiver **100**. The claims include a 1014-Hz laser frequency and a 108-Hz IF beam modulation.—AJC

7,039,208

43.66.Ts BEHIND-THE-EAR HOUSING FUNCTIONING AS A SWITCH

Josef Wagner and Daniel Oertli, assignors to Phonak AG
2 May 2006 (Class 381/322); filed 5 August 2002

To make manual switch activation easier for persons with limited dexterity, a pressure-sensitive, elastic, or bendable section in the center or end of a behind-the-ear hearing aid case is connected to an internal switching
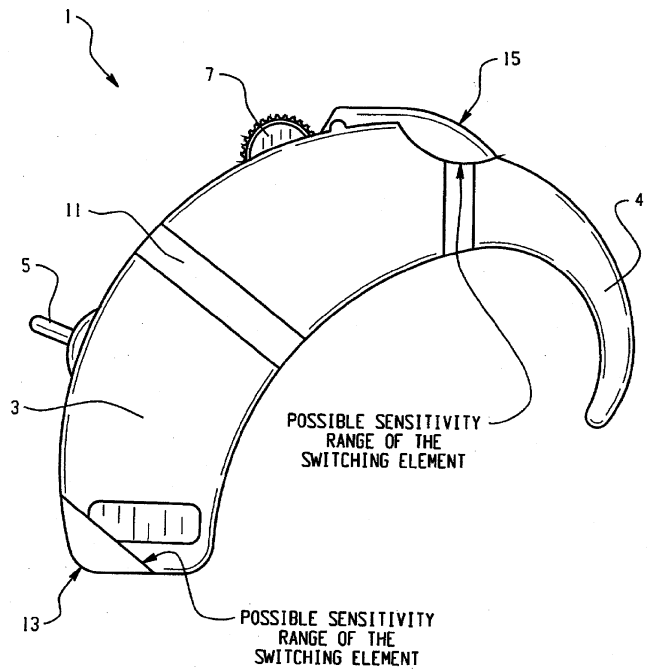
circuit element. Bending motion, compression, or deformation of the pressure-sensitive section triggers the switching operation.—DAP

7,039,209

43.66.Ts COVER DEVICE FOR HEARING AIDS

Christian Schmitt, assignor to Siemens Audiologische Technik GmbH
2 May 2006 (Class 381/322); filed in Germany 27 March 2002

The space required for a electronics/battery module is reduced (1) by using an Ω-shaped wire-spring-like element as an elastic hinge pin to pivot the battery drawer and (2) by removing the electronics portion from the
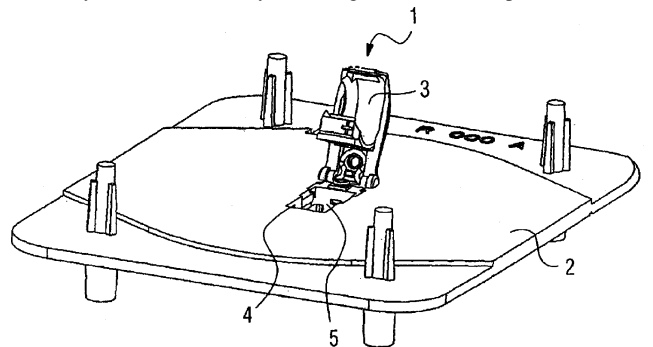
housing through the same opening used by the battery compartment after the battery compartment and wire spring are removed.—DAP

7,039,466

43.66.Ts SPATIAL DECIMATION STIMULATION IN AN IMPLANTABLE NEURAL STIMULATOR, SUCH AS A COCHLEAR IMPLANT

William Vanbrooks Harrison and Michael A. Faltys, assignors to Advanced Bionics Corporation
2 May 2006 (Class 607/56); filed 28 April 2004

Spontaneous neural activity is restored by inducing stochastic resonance via high-rate electrical stimulation. Power required from the battery is kept low via reducing the stimulus rate tonotopically by determining
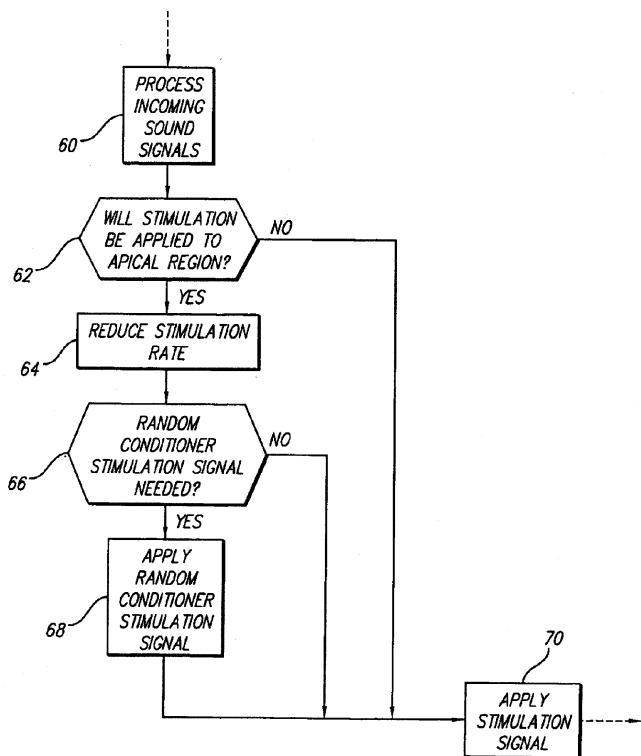
whether synchronous neural firings occur in the low-frequency region, for which high-rate pulses produce little benefit.—DAP

7,043,037

43.66.Ts HEARING AID HAVING ACOUSTICAL FEEDBACK PROTECTION

George Jay Lichtblau, Ridgefield, Connecticut
9 May 2006 (Class 381/317); filed 16 January 2004

Undesirable signals created internal to the hearing aid via acoustical or mechanical feedback are distinguished from desirable ambient signals and
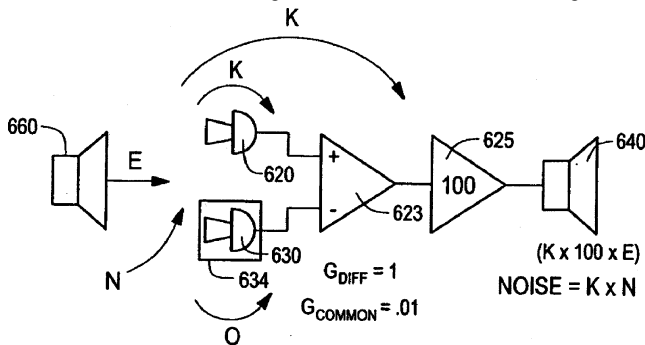
attenuated by taking the difference between two microphones' output signals. Both microphones are attached to the hearing aid housing, but only one microphone is positioned to pick up the desired ambient signals.—DAP

7,043,038

43.66.Ts IN-THE-EAR HEARING DEVICE

Herbert Bächler _et al._, assignors to Phonak AG
9 May 2006 (Class 381/322); filed 24 July 2002

An elastic textile material covers the rigid outer surface of an ITE hearing aid module and is adaptable to changes in the shape of the ear canal
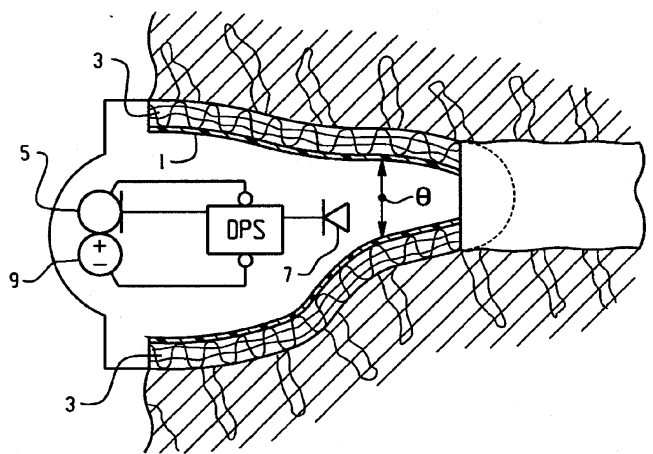
of the wearer during talking or chewing. Fitting comfort relative to that of an ITE hearing aid with a hard outer shell is said to be improved.—DAP

7,043,039

43.66.Ts HOUSING FOR A HEARING AID OR HEARING DEVICE RESPECTIVELY

Bruno Gabathuler, assignor to Phonak AG
9 May 2006 (Class 381/322); filed in the European Patent Office 21 November 2003

To prevent damaging force from being transferred to wires and components in an electronics module of a modular hearing device, the module has a battery compartment to hold the battery with a cap pivotable over a
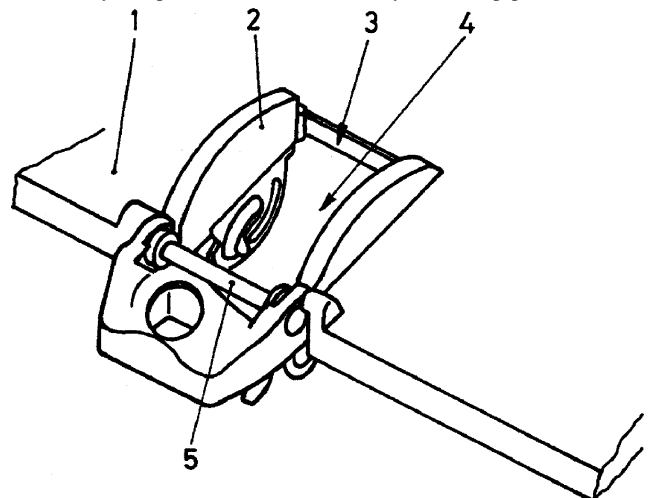
hinge pin. A bushing on each end of the hinge pin slips into receptacles in the faceplate of the modular device.—DAP

7,043,040

43.66.Ts HEARING AID APPARATUS

Patrick Westerkull, assignor to P&B Research AB
9 May 2006 (Class 381/326); filed in Sweden 21 June 2001

For hearing impaired persons with unilateral hearing loss, a bone-anchored hearing aid is implanted in the skull on the deaf side of the patient,

2411   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2411
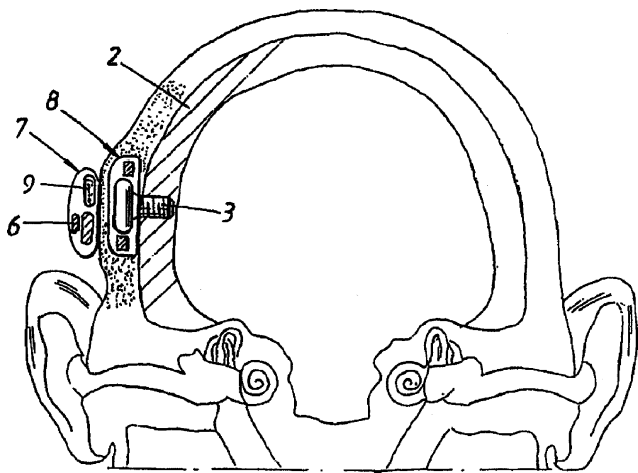
forming a "CROS" hearing aid device. High-frequency vibrations, in particular, are conducted through the skull bone to the inner ear of the better ear.—DAP

7,043,041

43.66.Ts INTEGRATED TELECOIL AMPLIFIER WITH SIGNAL PROCESSING

Miroslav hacek over (S)vajda and Alwin Fransen, assignors to SonionMicrotronic Nederland B.V.
9 May 2006 (Class 381/331); filed 3 October 2001

The frequency response of an integrated hearing-aid telecoil amplifier may be shaped with internal filtering for various input sources. The frequency shaping is independent of telecoil properties and may be used with
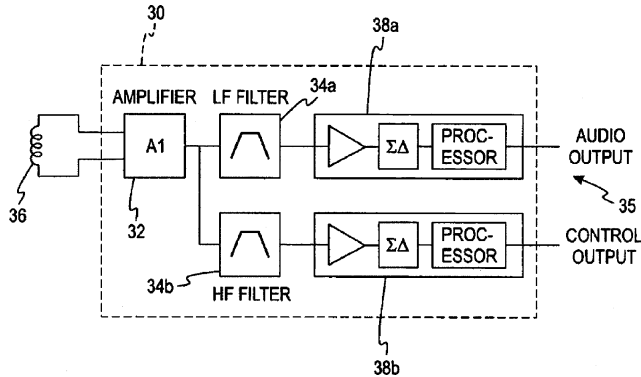
two signal paths for both audio frequency and radio frequency inputs. Radio frequency applications include programming and controlling the hearing aid.—DAP

7,043,303

43.66.Ts ENHANCED METHODS FOR DETERMINING ISO-LOUDNESS CONTOURS FOR FITTING COCHLEAR IMPLANT SOUND PROCESSORS

Edward H. Overstreet, assignor to Advanced Bionics Corporation
9 May 2006 (Class 607/57); filed 25 August 2003

Automatic self-programming for cochlear implant patients is made possible by predicting iso-loudness contours from iso-neural response contours. That is, minimum threshold level contours are predicted from measured moderately loud (M) level contours, which eliminates the need for patient responses. M contours are determined by setting a volume adjust-

ment of the signal in one channel until it is judged equal in loudness to that of a signal in a second channel and repeating this procedure for all channels.—DAP

7,050,966

43.66.Ts SOUND INTELLIGIBILITY ENHANCEMENT USING A PSYCHOACOUSTIC MODEL AND AN OVERSAMPLED FILTERBANK

Todd Schneider et al., assignors to AMI Semiconductor, Incorporated
23 May 2006 (Class 704/200.1); filed in Canada 7 August 2001

A desired signal is made audible over an undesired signal in high-noise environments as follows: the level of the undesired signal is measured either outside of the headset in the ambient environment, or under the headset, or in the ear canal of the wearer. After noise estimation and signal detection,
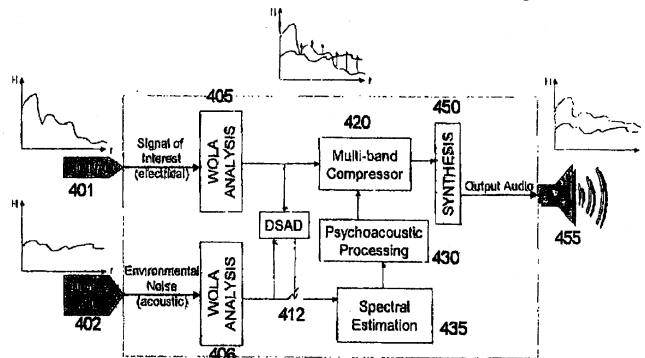
the level of the estimated desired signal is adaptively adjusted relative to the estimated level of the undesired signal. The dynamic range of the desired signal is mapped over the frequency range into the available dynamic range of the wearer to prevent excessively loud levels.—DAP

7,039,195

43.66.Vt EAR TERMINAL

Jarle Svean et al., assignors to Nacre AS
2 May 2006 (Class 381/71.6); filed 1 September 2000

A self-contained, idealized, in-the-ear noise protection/communication/dosimeter device contains both ear canal and environmental microphones, a speaker, and a ventilation channel with a pressure-release valve and a bypass path. Active noise cancellation is performed on the residual ear
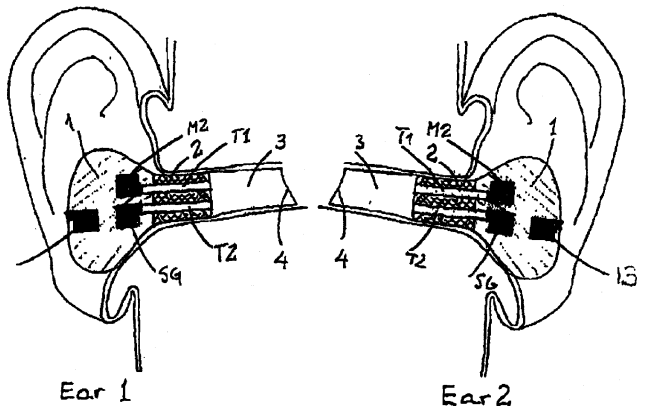
canal signal. The output of the device is transmitted via wireless means to a similar device in the opposite ear of the wearer. No mention is made about how to fit and power all of this into an in-the-ear device.—DAP

43.66.Yw HEARING TEST APPARATUS AND METHOD HAVING AUTOMATIC STARTING FUNCTIONALITY

Steven J. Iseberg *et al.*, assignors to Etymotic Research, Incorporated
23 May 2006 (Class 381/60); filed 2 March 2000

This is a device to analyze the distortion product otoacoustic emissions (DPOAEs) generated by the ear in order to determine middle ear function. In general a DPOAE test device generates and emits two audible tones (i.e., test signals) at different frequencies into a subject's ear canal. A healthy ear will produce, in response to the two audible tones, a response signal having a frequency that is a combination of the frequencies of the two audible tones. In this version of the DPOAE test device, the device analyzes the responses to determine whether the testing probe has been properly placed in the ear canal. The device may determine, e.g., whether the probe is stable, whether it is properly sealed in the ear canal, whether the resulting volume of the ear is acceptable, and/or whether the stimulus delivery is blocked in any way. If the device indicates that the testing probe has been properly placed in the ear canal, the device automatically starts a DPOAE test without requiring any operator input.—DRR

6,971,993

43.70.Dn METHOD FOR UTILIZING ORAL MOVEMENT AND RELATED EVENTS

Samuel G. Fletcher, assignor to Logometrix Corporation
6 December 2005 (Class 600/587); filed 14 November 2001

Described as a device to help a speaker improve the articulation of speech, this is a computer station with one or two attached palatography sensor plates. When in use, the speaker's areas of tongue contact with the palate sensor are displayed, together with an optional second display of
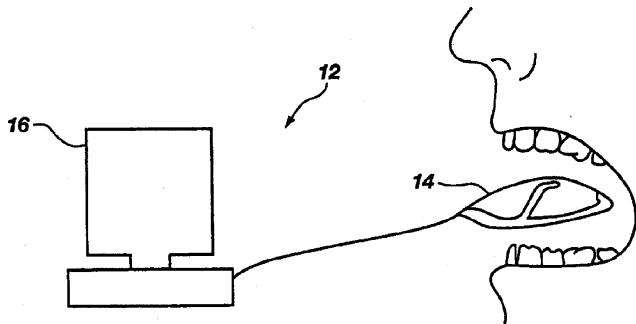
either a second speaker or a computer-generated reference palate-contact model. The patent includes a brief discussion of the patterns of palatal contact that might be expected during normal speech.—DLR

7,043,424

43.72.Ar PITCH MARK DETERMINATION USING A FUNDAMENTAL FREQUENCY BASED ADAPTABLE FILTER

Jau-Hung Chen and Yung-An Kao, assignors to Industrial Technology Research Institute
9 May 2006 (Class 704/207); filed in Taiwan 14 December 2001

This patent proposes to pinpoint the fundamental period of voiced speech by means of filtered waveform zero-crossing points. The signal must first go through a tight adaptive bandpass filter that closely tracks the fundamental frequency, so that all harmonics are filtered out. Of course, this necessitates actually tracking the fundamental frequency by some unmentioned means, and the whole approach seems a tad circular—one senses some labor is being needlessly repeated somewhere by a scheme that tracks pitch in order to tightly filter so that the pitch can be tracked.—SAF

43.72.Bs DISTRIBUTED SPEECH RECOGNITION WITH CODEC PARAMETERS

Bhiksha Raj *et al.*, assignors to Mitsubishi Electric Research Laboratories, Incorporated
23 May 2006 (Class 704/219); filed 27 November 2001

An encoded bitstream is first decoded to produce linear predictive coding (LPC) filter parameters and a residual signal, which are then combined using linear discrimination analysis to form speech recognition
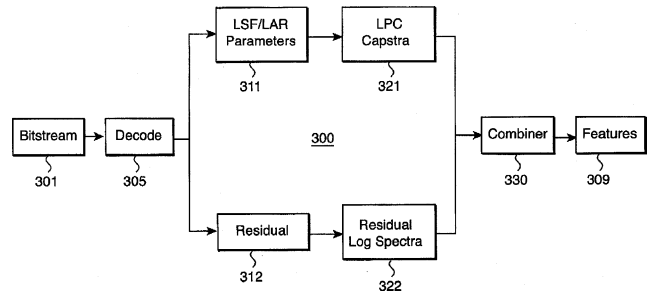
features. The LPC parameters represent the important short-term spectral characteristics and the residual signal models the long-term information such as pitch and perceptual quality. The method may be applied to various coding methods including LPC, GSM, and CELP.—DAP

7,043,030

43.72.Dv NOISE SUPPRESSION DEVICE

Satoru Furuta, assignor to Mitsubishi Denki Kabushiki Kaisha
9 May 2006 (Class 381/94.1); filed in Japan 9 June 1999

This extensive patent includes 15 embodiments and 48 claims, all surrounding a noise filtration scheme that is supposedly able to clean noise from a signal in a way that avoids numerous typical problems, including
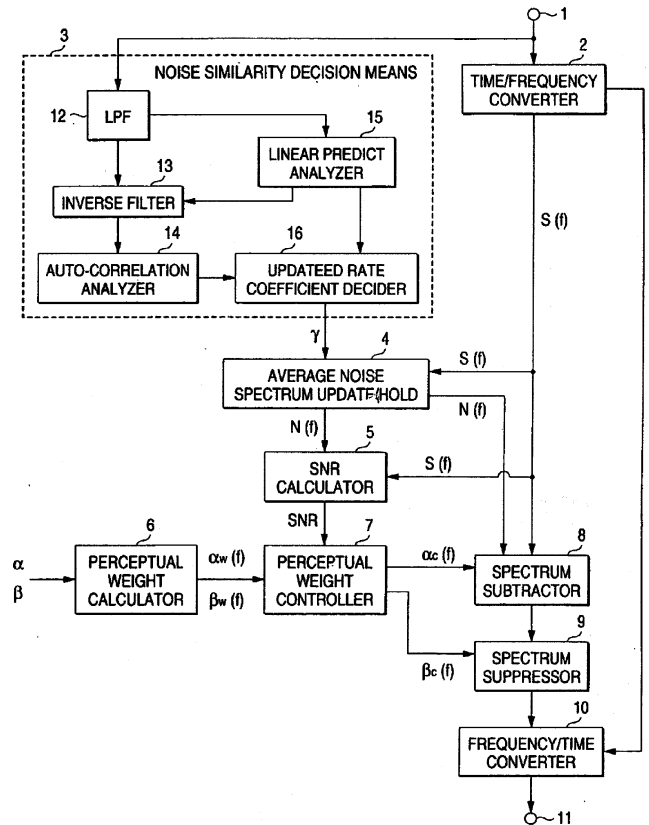
introduction of musical noise and excessive reduction of signal amplitude. The central idea seems to be the use of two or three ''perceptual weights'' that adapt to signal/noise ratio in a complicated spectral subtraction algorithm. How it works is presented in detail; why it works is still mysterious.—SAF

6,970,819

43.72.Ja SPEECH SYNTHESIS DEVICE

Yukio Tabei, assignor to Oki Electric Industry Company, Limited
29 November 2005 (Class 704/256); filed in Japan 17 March 2000

The patent describes a strategy to be used for assigning durations to vowels and consonant closures for use by a parametric speech synthesizer. Consonants are divided into classes, applying duration parameters to continuant portions and/or closure portions as appropriate. The actual method used to determine the durations is not described other than by reference in the patent text to two Japanese patent applications. However, these publications are not listed as references on the front page.—DLR

6,970,820

43.72.Ja VOICE PERSONALIZATION OF SPEECH SYNTHESIZER

Jean-Claude Junqua et al., assignors to Matsushita Electric Industrial Company, Limited
29 November 2005 (Class 704/258); filed 26 February 2001

A system is described for generating controls for a parametric speech synthesizer, with the intent that the synthesized output would closely approach the speech characteristics of a particular human speaker. This would be done by constructing the control parameters from two separate influences, those due to the speech content and those due to speaker characteristics. The novel element here is the construction of a speaker eigenspace based on a fairly short sample of representative speech. The eigenspace model can then be consulted to determine values for parameters not explicitly present in the speech sample. Construction of the eigenspace is covered in some detail in the patent text, but without mathematics.—DLR

6,975,987

43.72.Ja DEVICE AND METHOD FOR SYNTHESIZING SPEECH

Seiichi Tenpaku and Toshio Hirai, assignors to Arcadia, Incorporated
13 December 2005 (Class 704/258); filed in Japan 6 October 1999

This patent presents a technique for improving the quality of speech synthesized by concatenation of stored waveform segments. In such a system, the stored segments will typically need to be modified in order for the synthesized speech pitch to have the proper contours. The patent presents a
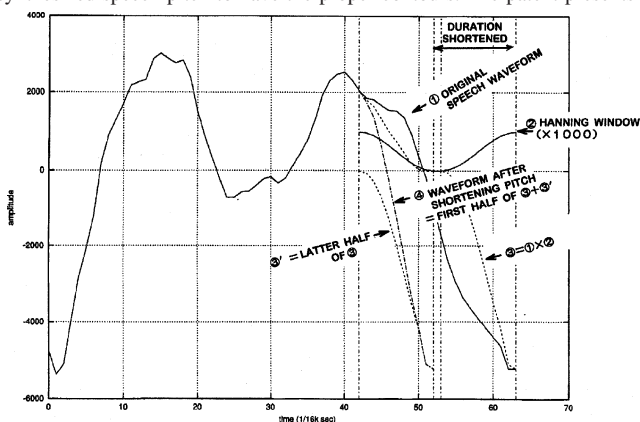
method of modeling the speech waveform so that a cut may be made for the purpose of pitch and/or duration modification while minimizing the effects of distortion of the speech spectral qualities.—DLR

6,975,988

43.72.Ja ELECTRONIC MAIL METHOD AND SYSTEM USING ASSOCIATED AUDIO AND VISUAL TECHNIQUES

Adam Roth, New York, New York et al.
13 December 2005 (Class 704/260); filed 10 November 2000

A technique is presented here for delivering an email message, in a multimedia form that may involve text and/or voice, as well as a facial image that appears to be voicing the words. The original message may have been in the form of either text or speech. At the receiving end, the facial image may be the actual speaker's image, the image of another person, or perhaps a cartoon image. A central aspect of the patented system is the ability to include emotional content into the message, which would not be conveyed by text alone. The patent describes different ways this goal may be achieved under the various combinations of modalities on the sending and receiving ends. What is still missing here is the yet unanswered but vital question of eye contact.—DLR

6,970,185

43.72.Ne METHOD AND APPARATUS FOR ENHANCING DIGITAL IMAGES WITH TEXTUAL EXPLANATIONS

Paul Steven Halverson, assignor to International Business Machines Corporation
29 November 2005 (Class 348/207.1); filed 31 January 2001

The basic scheme patented here is that a digital camera would provide a speech recognition service, allowing the photographer to speak a phrase to be digitally attached to the photo. Since a reasonably thorough and accurate recognition system would exceed the camera's ''on board'' resources, the
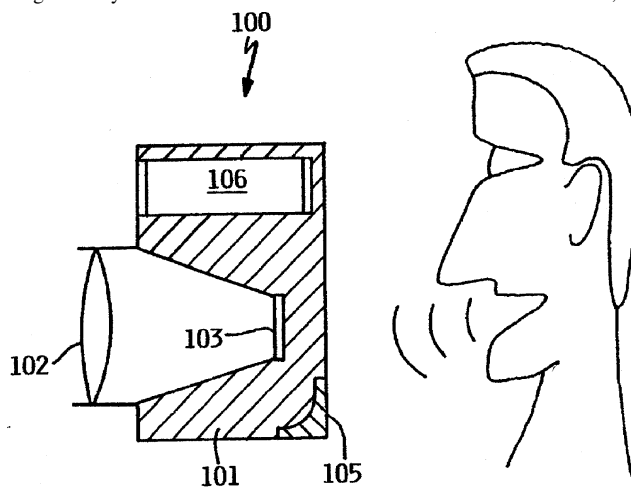
first step would involve a simpler version of audio analysis, producing a sort of phonetic script. Later, when the photo is downloaded to a computer, a more complete recognition system would finish the conversion of the phonetic script to a phrase in proper English text.—DLR

6,975,983

43.72.Ne NATURAL LANGUAGE INPUT METHOD AND APPARATUS

Nicholas David Fortescue and Robert Alexander Keiller, assignors to Canon Kabushiki Kaisha
13 December 2005 (Class 704/9); filed in the United Kingdom 29 October 1999

This patent covers the use of a natural language grammar as the language reference needed in a speech recognition system and, in particular, it describes how the natural language grammar may be modified so as to specify points in time where the speaker might pause, thus improving the recognition accuracy. The application involves the entry of speech data into a computer while the user also operates a mouse or keyboard. This might often involve the overlapping of speech with computer events, i.e., clicks or keystrokes, using these as a way of pointing to an item that is also mentioned verbally. The recognition grammar would use this dual-mode information to correctly identify the referent item.—DLR

7,039,589

43.72.Ne INTERACTIVE MULTIMEDIA BOOK

Charles Lamont Whitham, Reston, Virginia
2 May 2006 (Class 704/270); filed 1 December 2003

A computer-based book, containing both audio/video and text, provides hands-on multimedia instruction in response to voiced commands. The material is presented in a manner that is suitable for use in a variety of environments such as do-it-yourself repair. A word used as a hyperlink to other related topics changes its visual attributes when spoken and takes effect with a minimum of user interaction required, just as if a cursor had been placed on it.—DAP

7,043,425

43.72.Ne MODEL ADAPTIVE APPARATUS AND MODEL ADAPTIVE METHOD, RECORDING MEDIUM, AND PATTERN RECOGNITION APPARATUS

Hongchang Pao, assignor to Sony Corporation
9 May 2006 (Class 704/211); filed in Japan 28 December 1999

This patent introduces an adaptive method for adaptation of nonspeech models in speech recognition tools. The inventors believe that fresh nonspeech models can dynamically model noise characteristics, which may increase performance of speech recognition systems. To adapt nonspeech models with environmental noise, the invention has employed linear and nonlinear weighted functions. The model's parameters are altered based on the spectrum of the nonspeech signal.—AAD

7,043,429

43.72.Ne SPEECH RECOGNITION WITH PLURAL CONFIDENCE MEASURES

Sen-Chia Chang et al., assignors to Industrial Technology Research Institute
9 May 2006 (Class 704/236); filed in Taiwan, Province of China 24 August 2001

A speech recognition system for multiple languages avoids the need to formulate a complete set of speech data banks and acoustic models for each language. Two speech recognizers and three preset thresholds are used with
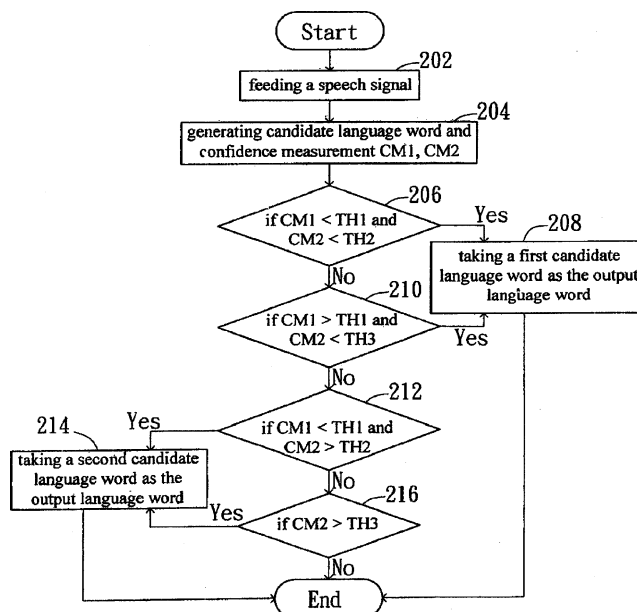
two confidence measures to determine whether a first or second candidate language word is best for the output language word.—DAP

7,043,430

43.72.Ne SYSTEM AND METHOD FOR SPEECH RECOGNITION USING TONAL MODELING

Grace Chung et al., assignors to Infotalk Corporation Limited
9 May 2006 (Class 704/251); filed 22 November 2000

A two-stage speech recognition methodology is proposed for application to tone languages such as Chinese. Stage one is, in essence, a standard speech recognition setup as would be applied to nontonal languages, but in which an effort is made to decode the speech syllabically. Stage two brings in the tone recognition scheme that concentrates on deciding which tone (pitch pattern) each syllable has been spoken with, and then the recognition scores from the two stages are combined to yield the predicted utterance.—SAF

7,047,192

43.72.Ne SIMULTANEOUS MULTI-USER REAL-TIME SPEECH RECOGNITION SYSTEM

Darrell A. Poirier, Quinebaug, Connecticut
16 May 2006 (Class 704/235); filed 27 June 2001

A speech recognition system creates a textual representation of a discussion by multiple speakers—a conference-to-text system. The system may include real-time language translation if multiple languages are used.—DAP

7,050,973

43.72.Ne SPEAKER RECOGNITION USING DYNAMIC TIME WARP TEMPLATE SPOTTING

Hagai Aronowitz, assignor to Intel Corporation
23 May 2006 (Class 704/246); filed 22 April 2002

Dynamic time warping involves aligning corresponding parts of an utterance against those of a target utterance. Spectral distances between coefficients are compared for spoken utterances against those for several

2415   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2415

```
                    200
           SPEAKER VERIFICATION
               PROCEDURE
                    |         202
           PERFORM TRAINING PROCEDURE
                    |         204
             RECEIVE TEST UTTERANCE
                    |         206
             SEGMENT TEST UTTERANCE
                   INTO FRAMES
                    |         208
           EXTRACT SPECTRAL COEFFICIENTS
                    |         210
            COMPUTE LOCAL DISTANCES
            BETWEEN TEST FRAMES AND
                 TARGET FRAMES
                    |         212
             COMPUTE ACCUMULATED
               DISTANCES FOR TEST
                AND TARGET FRAMES
                    |         214
            IDENTIFY TEST TEMPLATE(S)
                    |         216
            CALCULATE DTW DISTANCES
                    |         218
             GRANT ACCESS WHEN
               TEMPLATES MATCH
```

stored training templates, independent of the position and duration of the spectral components of the signal. Accumulated distances are used to determine endpoints of the utterances. The method is said to improve the accuracy of speaker identification for wireless devices used in noisy environments, while reducing required processing power.—DAP

7,050,974

43.72.Ne ENVIRONMENT ADAPTATION FOR SPEECH RECOGNITION IN A SPEECH COMMUNICATION SYSTEM

Yasuhiro Komori and Masayuki Yamada, assignors to Canon Kabushiki Kaisha
23 May 2006 (Class 704/256); filed in Japan 14 September 1999

This patent is mainly related to a remote speech recognition system. The system sends speech information to a remote host server through a communication network where recognition is performed. Environmental factors such as room acoustics, noise level, and communication-channel noise are calculated at the input terminal and sent to the server. In order to improve the performance of recognition, the speech models may be updated in the server upon noticing significant changes in the environmental factors.—AAD

7,039,587

43.72.Pf SPEAKER IDENTIFICATION EMPLOYING A CONFIDENCE MEASURE THAT USES STATISTICAL PROPERTIES OF *N*-BEST LISTS

Upendra V. Chaudhari *et al.*, assignors to International Business Machines Corporation
2 May 2006 (Class 704/246); filed 4 January 2002

A problem with typical statistical modeling approaches to speaker identification is identified as the inability to determine whether a particular identification result is confident or inconclusive. This patent puts forth techniques for statistically modeling the *N*-best list of candidate speakers that is output from a typical speaker identification scheme, training such models, and using the information provided to determine the confidence in an identification result.—SAF

7,038,117

43.75.Tv ELECTRONIC PERCUSSION INSTRUMENT AND VIBRATION DETECTION APPARATUS

Kiyoshi Yoshino, assignor to Roland Corporation
2 May 2006 (Class 84/411 R); filed in Japan 17 December 2002

Electronic drum pads not only need to sense the membrane, but also vibration on the frame (such as "rim shots"). As shown, the membrane **5** is

2416   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2416
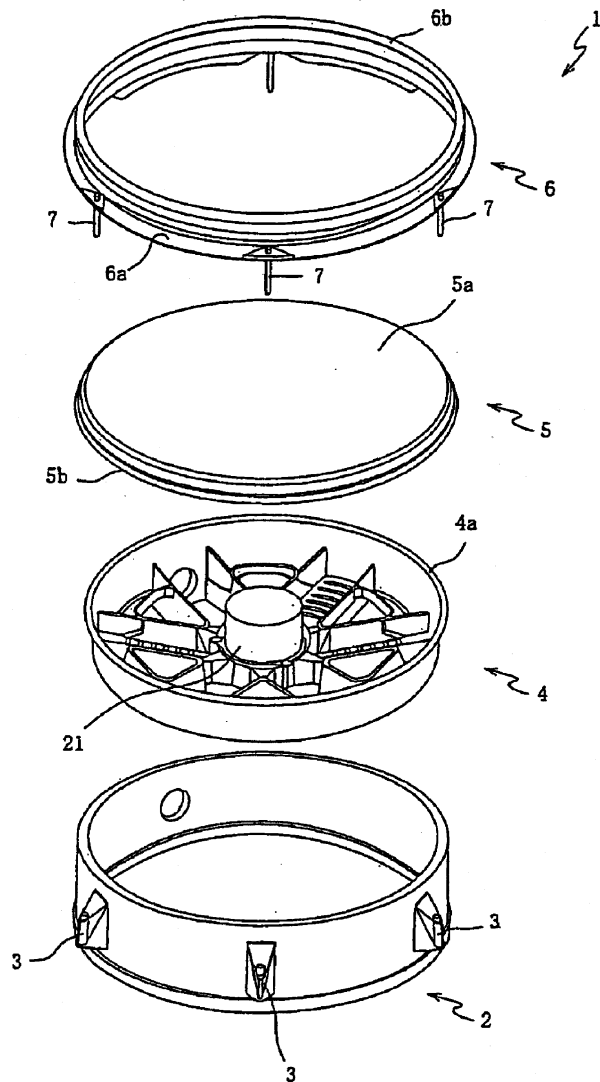
attached to the sensor **21**. Underneath **21** is another sensor affixed to the frame, shown inside, for detecting frame/rim hits. The overall design is clear and straightforward.—MK

7,038,119

43.75.Wx DYNAMIC CONTROL OF PROCESSING LOAD IN A WAVETABLE SYNTHESIZER

Andrej Petef, assignor to Telefonaktiebolaget L M Ericsson (publ)
2 May 2006 (Class 84/604); filed 18 July 2003

It is well known that the fixed sampling rate will impact the maximum number of voices in a synthesizer. A significant part of the computational load is devoted to the interpolation of ASDR (attack, sustain, decay, release) parameters. Naturally, the better the interpolation, the better the fidelity. This inventor proposes dynamically trading interpolator quality for the number of
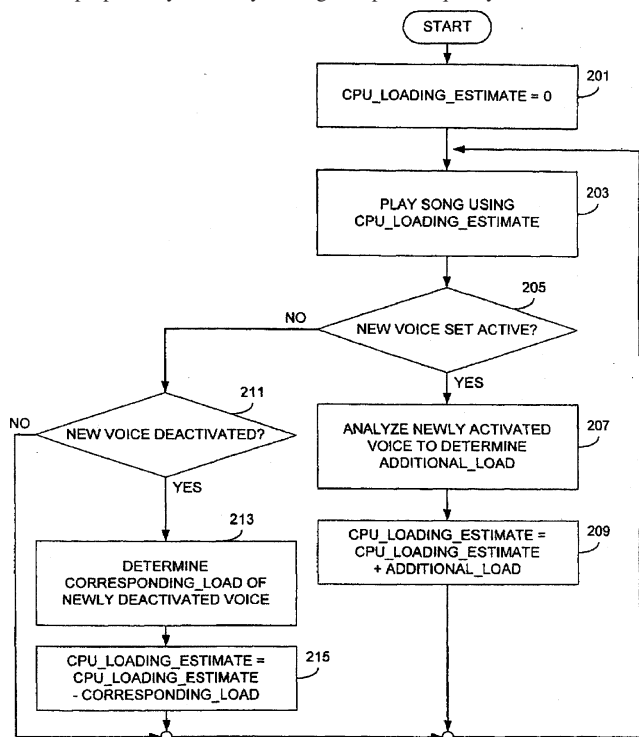
voices. Of course the distortion will increase as the number of voices in-ceases due to the decrease in interpolator quality. The inventor argues that the added voices will introduce additional masking effects, thereby reducing the audibility of the interpolator artifacts.—MK

7,039,478

43.75.Wx METHOD AND SYSTEM FOR AUTHORING A SOUNDSCAPE FOR A MEDIA APPLICATION

Eric Ledoux *et al.*, assignors to Microsoft Corporation
2 May 2006 (Class 700/94); filed 16 June 2004

As shown, this software can be divided into four stages: In stage 1, a flow graph is created. Stage 2 creates a graphic "control panel" as a graphi-cal user interface (GUI) while stage 3 assigns actions. All of this can be previewed before use (stage 4). Given this disclosure, you would think that
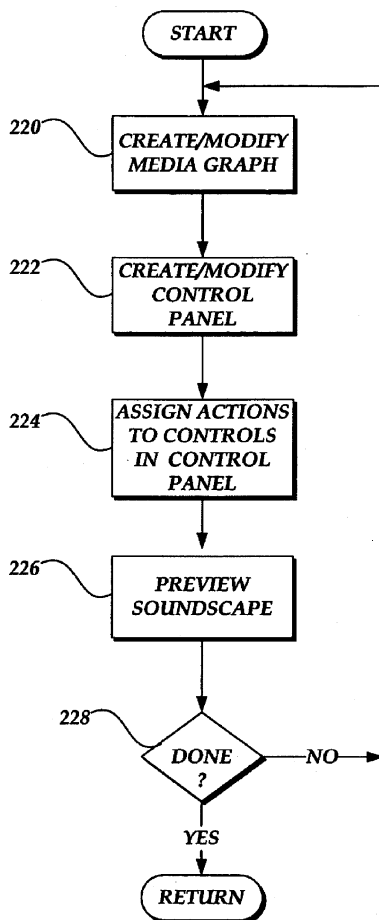
GUIs for sound did not exist before 1993. But is not this just a virtual sound scripting language? Where are references to Puckette's Max? Where is Tcl/Tk in all this?—MK

7,041,059

43.80.Qf 3D ULTRASOUND-BASED INSTRUMENT FOR NON-INVASIVE MEASUREMENT OF AMNIOTIC FLUID VOLUME

Vikram Chalana *et al.*, assignors to Diagnostic Ultrasound Corporation
9 May 2006 (Class 600/437); filed 20 May 2003

Disclosure is made of a hand-held 3-D ultrasound instrument for au-tomatically and noninvasively measuring amniotic fluid volume in the
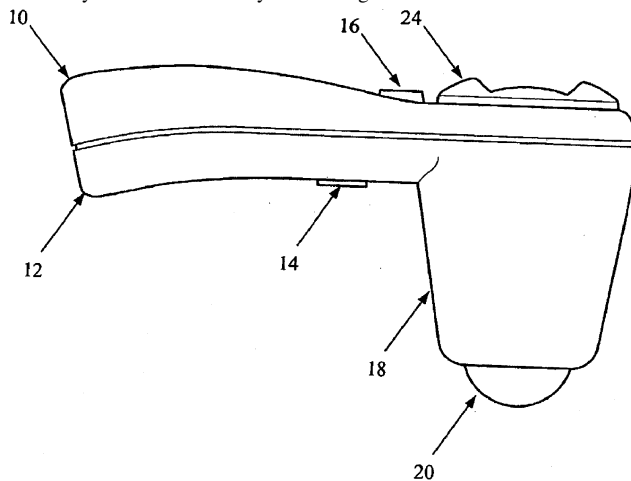
uterus. Through the use of a 2-D image-processing algorithm, this instrument provides the medical specialist with automatic feedback information as to where to acquire the 3-D image set. The user obtains one or more 3-D data sets covering all of the amniotic volume in the uterus, and this data is then processed using an optional 3-D algorithm to yield the total amniotic fluid volume corrected for any contributions from the fetal head volume.—DRR

7,044,913

43.80.Qf ULTRASONIC DIAGNOSIS APPARATUS

Eiichi Shiki, assignor to Kabushiki Kaisha Toshiba
16 May 2006 (Class 600/454); filed in Japan 15 June 2001

The goal of this ultrasonic apparatus is to effectively display dynamic blood-flow states in a patient being examined, and, in particular, to display pulsation of the blood flow, and to provide a 3-D view of pulsatile flows of blood. The apparatus consists of an ultrasound probe, a transmitter (that includes a transmitting pulse generator and a transmitting beamformer), a receiver (that includes a preamplifier and a receiving beamformer), a color-flow mapping processor, which incorporates a moving-element signal
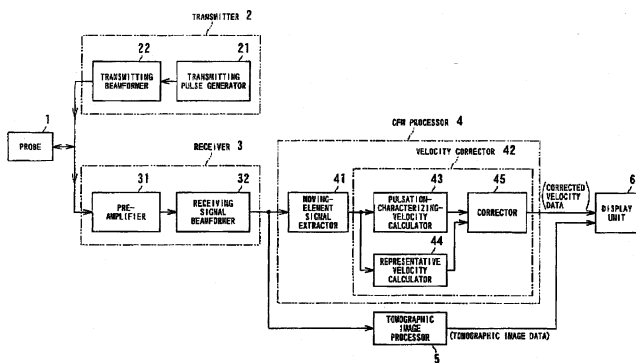
extractor, and a velocity corrector, a tomographic image processor, and a display unit. The device scans an area of interest by transmitting and receiving an ultrasound pulse to and from the patient. The velocity corrector calculates a pulsation-characterizing velocity, a representative (i.e., a reference) velocity, and corrects the velocities of the moving element on the basis of the standard velocity. The corrected velocity data are visualized on the display unit.—DRR

7,048,690

43.80.Qf PRECISION ULTRASOUND MEASUREMENT FOR INTRAOCULAR LENS PLACEMENT

Jackson D. Coleman *et al.*, assignors to Cornell Research Foundation, Incorporated
23 May 2006 (Class 600/452); filed 19 March 2002

The patent relates to a method of collecting, processing, and displaying data that are generated during scanning of the eye utilizing high-frequency ultrasound apparatus and associated devices. The geometry and/or topography of the relevant portions of the eye are depicted in preparation for lens
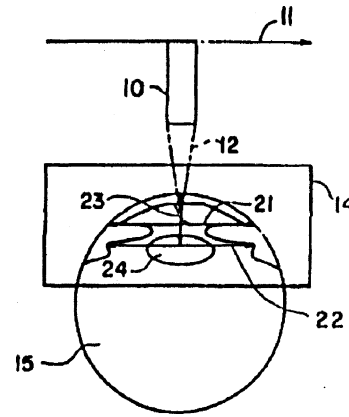
implantation and/or lens replacement in the case of cataract removal. Visual displays of the eye geometry and/or topography are obtained from data generated during a number of angularly spaced ultrasound scans taken across a meridional coronal section or a marginal section of the anterior surface of the eye.—DRR

7,052,462

43.80.Qf ULTRASONIC PROBE AND ULTRASONIC DIAGNOSTIC EQUIPMENT

Hiroshi Fukuda and Masayoshi Omura, assignors to Olympus Corporation
30 May 2006 (Class 600/445); filed 24 October 2002

The object of this device is to provide an ultrasonic probe in which ultrasonic attenuation at high frequencies is small and the decrease of the capacity of an acoustic medium, due to volatilization, can be reduced. A flexible shaft is inserted through a sheath constituting an insertion portion of an ultrasonic probe connected to an ultrasonic observation platform. An ultrasonic transducer, driven to rotate, is attached at the distal end. The attenuation at high frequencies can be reduced through the use of a hydrocarbon-based oil having a kinematic viscosity of 20 mm^2/s as an acoustic medium surrounding the ultrasonic probe in the acoustic window. The decrease in the capacity of the acoustic medium due to volatilization is also minimized.—DRR

7,052,467

43.80.Qf STETHOSCOPIC SYSTEMS AND METHODS

Shawn C. D. Johnson, Winchester and David S. Geller, Lexington, both of Massachusetts
30 May 2006 (Class 600/528); filed 4 October 2002

This is a device for communicating stethoscope sounds over a telephone system. The device consists of a chest piece, an amplifier, and a mixer compatible with a conventional voice-quality telephone system. Electrical signals are generated that are representative of the physiological sounds, such as breathing sounds, heart sounds, and bowel sounds. The electrical signals can be combined with voice communications at the remote site and
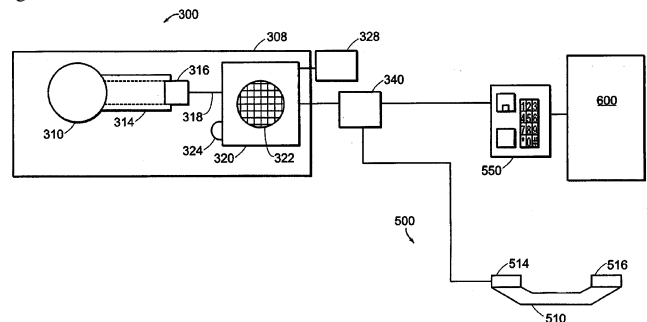
transmitted over the telephone system. A system and method is also provided for diagnosis by a trained health-care professional with the patient being located at a remote site. In a second embodiment, a system and method are also provided for training students to interpret stethoscope sounds.—DRR

7,044,737

43.80.Sh ULTRASOUND ORAL HYGIENE AND THERAPEUTIC DEVICE

Liang Fu, Potomac, Maryland
16 May 2006 (Class 433/119); filed 5 March 2004

An ultrasound therapeutic device uses a combination of ultrasonic transducer array and a filling means to control the ultrasound radiation to each part of the oral cavity that is being treated for cleaning and therapy. Every surface of the oral cavity, including the interdental and gingival areas, is cleaned and treated for therapy simultaneously, uniformly, and with
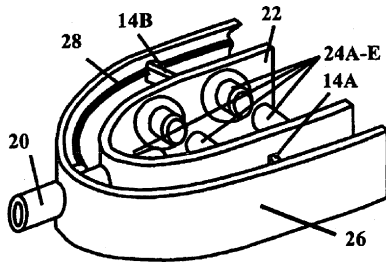
optimal amount of ultrasound radiation. It is also asserted that every part of the oral cavity is protected from an overdose of ultrasound radiation through limited ultrasound action zones for specific exposure times. The soft, resilient filling element sits snugly into a patient's oral cavity and it serves to confine the cleaning solution and the ultrasound action.—DRR

7,044,960

43.80.Sh METHOD AND APPARATUS FOR PROVIDING NON-INVASIVE ULTRASOUND HEATING OF THE PREOPTIC ANTERIOR HYPOTHALAMUS

Marc E. Voorhees and Gary A. Carson, assignors to Medivance Incorporated
16 May 2006 (Class 607/96); filed 17 September 2003

Among neuroprotectant approaches to treating strokes is the induction of mild hypothermia, which has been shown to inhibit the chemical cascade that causes secondary cellular death after an ischemic event and to provide broad neuroprotection. The patent covers a method and system to induce mild hypothermia in a patient through controlled heating of the preoptic anterior hypothalamus (POAH) in conjunction with cooling of the patient's body. An ultrasound transducer is positioned extracorporeally to the patient's skull in order to transmit ultrasound energy to the POAH. This heats the POAH to inhibit thermoregulatory responses so that the patient's core
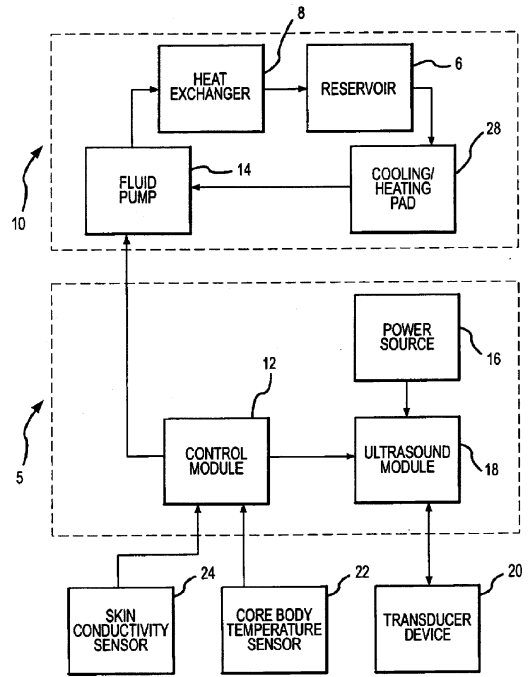
body temperature may more effectively be cooled. Feedback sensors may be strategically placed at various locations on the patient in order to monitor the core temperature. A control subsystem controls the amount of heat the POAH receives and the amount of cooling the body receives on the basis of feedback signals from the sensors.—DRR

7,052,506

43.80.Sh ULTRASONIC SURGICAL TOOL

Michael John Radley Young and Stephen Michael Radley Young, both of South Devon, the United Kingdom
30 May 2006 (Class 606/169); filed in the United Kingdom 17 September 1999

The goal of this surgical device is to provide a surgical tool with a disposable shroud system that can be quickly and easily detached from the apparatus and replaced by a replacement shroud. The tool consists of an ultrasonic vibration generator with a waveguide at one end. The distal end of the waveguide is provided with cutting and/or coagulating means. A shroud
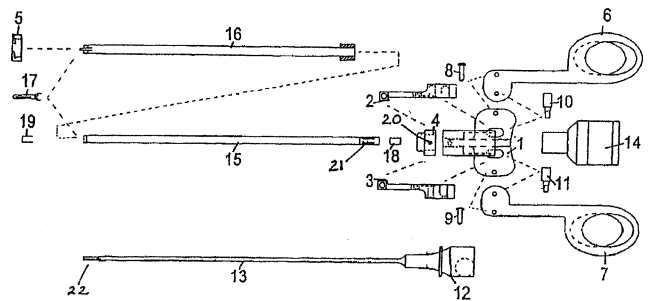
is adapted to surround and isolate the waveguide and an actuating rod surrounds the rod or vice versa. The shroud and actuating rod are detachably connected at the proximal ends to the generator.—DRR

2419   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2419

7,044,912

43.80.Vj DIAGNOSTIC MEDICAL ULTRASOUND SYSTEM HAVING METHOD AND APPARATUS FOR STORING AND RETRIEVING 3D AND 4D DATA SETS

Sundar G. Babu *et al.*, **assignors to Siemens Medical Solutions USA Incorporated**
16 May 2006 (Class 600/437); filed 28 August 2003

User-defined viewing parameters are associated with image data to permit manipulation of the image data while the integrity of the underlying image data is maintained.—RCW

7,047,063

43.80.Vj TISSUE SITE MARKERS FOR IN VIVO IMAGING

Fred H. Burbank *et al.*, **assignors to SenoRx, Incorporated**
16 May 2006 (Class 600/431); filed 10 September 2003

These biopsy site markers are small but have a high ultrasound reflectivity to show the location of a biopsy cavity without obscuring diagnostic features in subsequent imaging studies.—RCW

7,052,460

43.80.Vj SYSTEM FOR PRODUCING AN ULTRASOUND IMAGE USING LINE-BASED IMAGE RECONSTRUCTION

Godwin Liu *et al.*, **assignors to VisualSonics Incorporated**
30 May 2006 (Class 600/443); filed 15 December 2003

Ultrasound energy with a center frequency of 20 MHz or higher is transmitted, received, and processed using scan lines to produce ultrasound images at an effective frame rate of 200 or more frames per second.—RCW

7,052,461

43.80.Vj METHOD AND SYSTEM FOR REGISTERING ULTRASOUND IMAGE IN THREE-DIMENSIONAL COORDINATE SYSTEM

Parker Willis, assignor to SciMed Life Systems, Incorporated
30 May 2006 (Class 600/443); filed 16 November 2004

Ultrasound images are acquired in one coordinate system and graphical data about the site of interest are acquired in another coordinate system. The location of the imaging transducer is determined in both coordinate
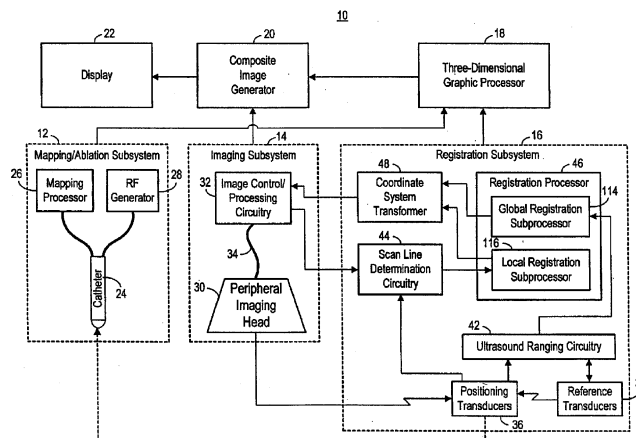
systems. The ultrasound images acquired in the first coordinate system are registered and displayed in the second coordinate system along with the graphical data.—RCW

2420   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Reviews of Acoustical Patents   2420

LETTERS TO THE EDITOR

An audio-visual corpus for speech perception and automatic speech recognition (L)

Martin Cooke[a)] and Jon Barker
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield, S1 4DP United Kingdom

Stuart Cunningham
Department of Human Communication Sciences, University of Sheffield, Sheffield, S1 4DP United Kingdom

Xu Shao
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield, S1 4DP United Kingdom

An audio-visual corpus has been collected to support the use of common material in speech perception and automatic speech recognition studies. The corpus consists of high-quality audio and video recordings of 1000 sentences spoken by each of 34 talkers. Sentences are simple, syntactically identical phrases such as "place green at B 4 now." Intelligibility tests using the audio signals suggest that the material is easily identifiable in quiet and low levels of stationary noise. The annotated corpus is available on the web for research use. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2229005]

I. INTRODUCTION

Understanding how humans process and interpret speech in adverse conditions is a major scientific challenge. Two distinct methods for modeling speech perception have been studied. The traditional approach has been to construct "macroscopic" models, which predict overall speech intelligibility in conditions of masking and reverberation. Models such as the articulation index (French and Steinberg, 1947), the speech transmission index (Steeneken and Houtgast, 1980), and the speech intelligibility index (ANSI S3.5, 1997) fall into this category. A more recent idea is to apply automatic speech recognition (ASR) technology to construct what might be called "microscopic" models of speech perception, which differ from macroscopic approaches in their additional capability to predict listeners' responses to individual tokens. Examples of microscopic models include Ghitza (1993), Ainsworth and Meyer (1994), Holube and Kollmeier (1996), and Cooke (2006).

Although microscopic modeling results have been promising, a serious barrier to further development of these models has been the lack of suitable speech material. Unlike speech perception studies, microscopic models require a large volume of speech material for training purposes. Many corpora for ASR exist, but the use of such corpora in speech perception testing is problematic. Speech material tends to be uncontrolled, phonetically unbalanced, or consists of tokens whose durations make them unsuitable for behavioral studies. By contrast, corpora used in perceptual studies tend to be too small or insufficiently varied for microscopic models of speech perception.

Previous models have attempted to explain the *auditory* perception of speech signals. However, speech production results in both acoustic and optical signals. It has become increasingly clear that the visual modality has a fundamental role in speech perception, and any full perceptual account needs to explain the complicated interactions between modalities (Rosenblum, 2002). Acoustically confusable phoneme pairs such as /m/ and /n/ can be disambiguated using visual cues. Automatic speech recognition systems can exploit these cues to improve audio-only recognition performance in both clean and noisy conditions (Potamianos *et al.*, 2003). Visual cues can also be used to separate speech from competing noise sources. One particularly interesting area of study in this respect is the audio-visual separation of simultaneous cochannel speech. Despite the clear importance of visual speech information, until now there have been no eas-

a)Electronic mail: m.cooke@dcs.shef.ac.uk

ily accessible corpora suitable for building multimodal models. However, recent advances in video compression technology, the rapidly falling cost of hard disk storage, the increasing capacity of optical storage media, and the increasing bandwidth of typical Internet connections, mean that storage and distribution are no longer a barrier.

These factors motivated the collection of an audio-visual corpus designed for both ASR-based and perceptual studies of speech processing. The form of the corpus was heavily influenced by the coordinate response measure (CRM) task (Moore, 1981; Bolia *et al.*, 2000), which consists of simple sentences of the form "READY <call sign> GO TO <color> <digit> NOW." CRM used 8 call signs, 4 colors, and 8 digits and each combination was spoken once by 8 talkers for a total of 2048 sentences. CRM is useful for studies of early processes in speech perception since it contains sentence length material, yet is devoid of high-level linguistic cues. The design of CRM makes it valuable in multitalker tasks (e.g., Brungart *et al.*, 2001) where listeners are asked to identify the color-digit combination spoken by the talker who provided a given call sign.

The new collection, which we call the Grid corpus, consists of sentences such as "place blue at F 9 now" of the form <command:4> <color:4> <preposition:4> <letter:25 > <digit:10> <adverb:4>," where the number of choices for each component is indicated. Grid extends CRM in a number of ways. The set of talkers is larger (34 rather than 8) and the number of sentences per talker is 1000 rather than 256, giving a total corpus size of 34 000 as opposed to 2048 sentences. Consequently, Grid contains greater variety and is large enough to meet the training requirements of ASR systems. Grid has an improved phonetic balance due to the use of alphabetic letters, which also presents listeners with a more difficult task than the four color options of CRM. Grid is more varied than CRM since the "filler" items (command, preposition, and adverb) are no longer static. This also prevents echo-like artifacts arising when two or more sentences with identical fillers are summed in, for example, experiments involving multiple simultaneous talkers. Finally, Grid provides speech video as well as audio, allowing the development of multimodal perceptual models.

While the primary motivation for the Grid corpus was to support the construction of microscopic, multimodal models of speech perception, it can also be used for conventional behavioral studies of audio and audio-visual speech perception. Similarly, Grid is valuable for ASR studies of speech in noise, the separation of speech from multitalker backgrounds, and audio-visual speech recognition and separation.

II. CORPUS

A. Sentence design

Each sentence consisted of a six word sequence of the form indicated in Table I. Of the six components, three—color, letter, and digit—were designated as "keywords". In the letter position, "w" was excluded since it is the only multisyllabic English alphabetic letter. "Zero" was used rather than "oh" or "nought" to avoid multiple pronunciation alternatives for orthographic "0." Each talker produced all combinations of the three keywords, leading to a total of

TABLE I. Sentence structure for the Grid corpus. Keywords are identified with asterisks.

command	color*	preposition	letter*	digit*	adverb
bin	blue	at	A–Z	1–9, zero	again
lay	green	by	excluding W		now
place	red	in			please
set	white	with			soon

1000 sentences per talker. The remaining components—command, preposition, and adverb—were "fillers." Four alternatives were available in each filler position. Filler words were chosen to create some variation in contexts for the neighboring key words. Different gross phonetic classes (nasal, vowel, fricative, plosive, liquid) were used as the initial or final sounds of filler words in each position.

B. Speaker population

The aim of speaker selection was to provide a sufficiently large number of speakers to allow users of the corpus to select subsets based on criteria such as intelligibility, homogeneity, and variety. Sixteen female and 18 male talkers contributed to the corpus. Participants were staff and students in the Departments of Computer Science and Human Communication Science at the University of Sheffield. Student participants were paid for their contribution. All spoke English as their first language. All but three participants had spent most of their lives in England and together encompassed a range of English accents. Two participants grew up in Scotland and one was born in Jamaica. Ages ranged from 18 to 49 years (mean: 27.4 years).

C. Collection

Audio-visual recordings were made in an IAC single-walled acoustically isolated booth. Speech material was collected from a single Bruel & Kjaer (B & K) type 4190 $\frac{1}{2}$ -in. microphone placed 30 cm in front of the talker. The signal was preamplified by a B & K Nexus model 2690 conditioning amplifier prior to digitization at 50 kHz by a Tucker-Davis Technologies System 3 RP2.1 processor. Collection of speech material was under computer control. Sentences were presented on a computer screen located outside the booth, and talkers had 3 s to produce each sentence. Talkers were instructed to speak in a natural style. To avoid overly careful and drawn-out utterances, they were asked to speak sufficiently quickly to fit into the 3-s time window. Talkers were allowed to repeat the sentence if they felt it necessary, either because of a mistake during production or if part of the utterance fell outside the 3-s window. As an aid, the captured waveform was displayed on the screen. In addition, talkers were asked to repeat the utterance if the captured waveform was judged by the software to be too quiet or too loud. Prior to saving, signals were scaled so that the maximum absolute value was unity, in order to optimize the use of the quantized amplitude range. Scale factors were stored to allow the normalization process to be reversed.

A simultaneous continuous video recording was made on to MiniDV tape using a Canon XM2 video camcorder. The camera was set up to capture full frames at 25 frames/s.

To avoid both noise and distraction from the video apparatus, the camera was placed at eye level outside the booth, abutting the booth window. Light sources were arranged to produce uniform illumination across the face, and the subject was seated in front of a plain blue background. To ease temporal alignment of the audio and visual signals, the camera took its audio input from the high-quality audio signal provided by the microphone in the booth.

Although talkers were allowed to repeat an utterance if they misread the prompt, they occasionally made errors without realizing they had done so. A semi-automatic screening procedure was employed to locate errors in the corpus. The screening process used an ASR system based on talker-dependent, whole-word hidden Markov models (HMM), trained separately for each talker. A "jack-knife" training procedure was used in which 80% of the talker's utterances were used for training and the remaining 20% recognized. This procedure was performed five times with a different subset of utterances so that the subset recognized was independent of the training set on each occasion. The ASR system produced word-level transcripts for each utterance, and errors were flagged if the recognition output differed from the sentence the talker was meant to have read. On average, 57 out of 1000 utterances were flagged per talker. Flagged utterances were checked by a listener and any sentences with errors were marked for re-recording. Talkers were recalled to perform the re-recording session, during which time their utterances were monitored over headphones. Talkers were asked to repeat any incorrectly produced utterances. In all, 640 utterances (an average of 19 per talker or 1.9% of the corpus) were re-recorded.

While the screening process guaranteed that many of the errors in the corpus were corrected, it is possible that some errors were not detected. For a spoken sentence containing an error to appear correct, the recognition system must have made a complementary error (i.e., an error which corrects the error made by the talker). However, since the error rates of both the talkers and recognizer are very low, the conjunction of complementary errors is extremely unlikely. Informal human screening of a subset of utterances led to an estimate of an error rate of not more than 0.1% (i.e., one error per 1000 utterances). Most of the errors detected involved misproduced filler items, so the number of sentences containing misproduced keywords is smaller still.

D. Postprocessing
1. Audio

Prior to further processing, audio signals were down-sampled to 25 kHz using the MATLAB *resample* routine. A subset of 136 utterances (four randomly chosen from each talker) was used to estimate the peak S/N according to the ITU P.56 standard (ITU-T, 1993). The peak S/N varied across talkers from 44 to 58 dB (mean=51 dB, s.d.=3.6 dB).

The talker-dependent HMM-based ASR systems used in the screening of speaker errors were employed to estimate the alignment between the word-level transcription and the utterance. In addition, phone-level transcriptions of each utterance were produced by forced alignment using the HVITE program in the HTK hidden Markov model toolkit (Young *et*

al., 1999). Pronunciations were taken from the British English Example Pronunciation dictionary (BEEP[1]). To bootstrap the initial set of HMMs, 60 sentences from one of the speakers were manually transcribed at the phone level.

2. Video

Unlike the audio collection, which was computer-controlled, the video data were collected continuously throughout the recording session, and thus contained both genuine Grid utterances as well as false starts, incorrect utterances, and other material. Consequently, it was necessary to extract video segments corresponding to the final end-pointed audio recordings. Utterance segments were first located approximately using a timestamp recorded by the software controlling the audio recording session. A precise location was then found by searching in this region for the 3-s period of the video file that best correlated with the 3 s of high-quality audio captured by the TDT processor. Correlations were performed using the smoothed energy envelope of the signals. Once the audio was precisely located, the corresponding 75-frame (i.e., 3-s) segment of video was extracted. The high-quality 50-kHz audio captured by the TDT processor was resampled to 44.1 kHz and used to replace the audio track of the video segment.

The DV format video was converted to MPEG-1 format using FFMPEG.[2] Two compression rates were used to produce both high and moderate quality versions of the video data. The high-quality video employed a bandwidth of 6 Mbits per s (comparable to DVD quality), while the moderate quality version used a bandwidth of 600 Kbits per s (a quality intermediate between a typical business-oriented videoconferencing system and VHS video). In both cases, the audio bit rate was set to 256 kbits per s.

III. AUDIO INTELLIGIBILITY TESTS

Twenty listeners with normal hearing heard independent sets of 100 sentences drawn at random from the corpus. All speech material had initial and trailing silence removed prior to presentation using utterance endpoints derived from the word alignments. Utterances were scaled to produce a presentation level of approximately 68 dB SPL and were presented diotically over Sennheiser HD250 headphones in the IAC booth. Listeners were asked to identify the color, letter, and digit spoken and entered their results using a conventional computer keyboard in which four of the nonletter/digit keys were marked with colored stickers. Those keys representing colors were activated immediately following the onset of each utterance. As soon as a color key was pressed, the 25 relevant letter keys were enabled, followed by the 10 digit keys. This approach allowed for rapid and accurate data entry: most listeners were able to identify a block of 100 utterances in 5–7 min. Listeners were familiarized with the stimuli and the task by identifying an independent practice set of 100 sentences prior to the main set.

Figure 1 (triangles) shows the mean scores and their standard errors across listeners. Unsurprisingly, fewer errors were made for colors (0.25% of the 2000 sentences) than for digits (0.7%) or letters (0.95%). At least one error occurred
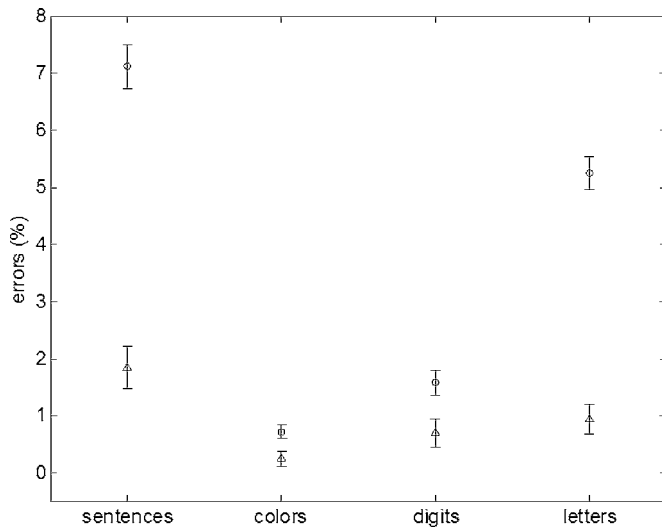
FIG. 1. Mean error rates across listeners for sentences, colors, digits, and letters. A sentence contained an error if one or more keywords were incorrectly identified. Triangles: clean sentence material; circles: sentences in speech-shaped noise. Error bars denote $+/-1$ standard errors.

FIG. 2. Error rates for colors, digits, letters, and per talker for the noise conditions. Horizontal lines indicate mean error rates in each category. Talker error rates are measured as percentages of utterances in which at least one keyword was misidentified.

in 1.85% (37 out of 2000) of sentences. These low error rates suggest that the speech material collected was of high intelligibility overall.

Insufficient errors were reported using clean speech material to allow a more detailed inspection of the intelligibility of individual keywords or talkers. To support such an analysis, the same 20 listeners heard three further independent sets of 100 utterances mixed with speech-shaped noise whose spectrum matched the long-term spectrum of the Grid corpus at three signal-to-noise ratios: 6, 4, and 2 dB, producing a total of 6000 responses. Figure 1 (circles) shows error rates for colors (0.7%), digits (1.6%), letters (5.2%), and wholly correct sentences (7.1%). Figure 2 depicts the distribution of errors by keyword and across talkers. While the color and number distributions are reasonably flat, certain letters are recognized significantly less well than others. Inspection of letter confusion matrices revealed that most of the /v/ errors were caused by misidentification as /b/, while /m/ and /n/ tokens were confused with each other.

A range of identification rates (defined as the percentage of utterances in which at least one keyword was misidentified) across the 34 contributing talkers was observed. In particular, listeners misidentified keywords in utterances by talkers 1 (19.8%), 20 (16.2%), and 33 (15.6%), while fewer than 2% of sentences spoken by talker 7 were misidentified by this listener group. However, most talkers produced errors rates of around 5%.

IV. SUMMARY

Grid, a large multitalker audio-visual sentence corpus, has been collected to support joint computational-behavioral studies in speech perception. Audio-only intelligibility tests suggest that the speech material is easily identified in quiet and low-noise conditions. Further tests of visual and audio-visual intelligibility are planned. The complete corpus and transcriptions are freely available for research use at the website <http://www.dcs.shef.ac.uk/spandh/gridcorpus>.

[1]Available at ftp://svr-ftp.eng.cam.ac.uk/pub/comp.speech/dictionaries/
[2]Available at http://ffmpeg.sourceforge.net/index.php

Ainsworth, W. A., and Meyer, G. F. (**1994**). "Recognition of plosive syllables in noise: Comparison of an auditory model with human performance," J. Acoust. Soc. Am. **96**, 687–694.

ANSI (**1997**). ANSI S3.5-1997, "American National Standard Methods for the Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **100**, 2527–2538.

Cooke, M. P. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

French, N., and Steinberg, J. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Ghitza, O. (**1993**). "Adequacy of auditory models to predict human internal representation of speech sounds," J. Acoust. Soc. Am. **93**, 2160–2171.

Holube, I., and Kollmeier, B. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," J. Acoust. Soc. Am. **100**, 1703–1716.

ITU-T (**1993**). "Objective measurement of active speech level," ITU-T Recommendation P. 56.

Moore, T. (**1981**). "Voice communication jamming research," in *AGARD Conference Proceedings 331: Aural Communication in Aviation*, Neuilly-Sur-Seine, France, 2:1–2:6.

Potamianos, G., Neti, C., Gravier, G., Garg, A., and Senior, A. W. (**2003**). "Recent advances in the automatic recognition of audiovisual speech," Proc. IEEE **91**, 1306–1326.

Rosenblum, L. D. (**2002**). "The perceptual basis for audiovisual integration," in *Proceedings International Conference on Spoken Language Processing*, 1461–1464.

Steeneken, H., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.

Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V., and Woodland, P. (**1999**). *The HTK Book 2.2* (Entropic, Cambridge).

Reflection and transmission of waves in composite porous media: A quantification of energy conversions involving slow waves

J. Germán Rubino[a] and Claudia L. Ravazzoli
Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque, S/N, (1900) La Plata, Argentina and C.O.N.I.C.E.T., Argentina

Juan E. Santos
Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque, S/N, (1900) La Plata, Argentina, and C.O.N.I.C.E.T., Argentina and Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, Indiana 47907-2067

This paper studies the reflection and transmission of plane elastic waves at interfaces in fluid-saturated poroviscoelastic media in which the solid matrix is composed of two weakly coupled solids. The analysis of this problem, not formally performed before, is based on a theory recently developed by some of the authors, which allows us to derive expressions for the reflection and transmission coefficients at a plane interface within this kind of media and their relationship with the energy flux (Umov-Poynting) vector. The results of the present derivation were applied to study the energy splitting that takes place when a plane fast compressional wave strikes obliquely an interface defined by a change in ice content within a sample of water saturated partially frozen sandstone. The numerical results show wave mode conversions from fast to slow compressional and shear waves, with maximum energy conversion on the order of 20% from fast to slow wave modes near the critical angle. This phenomenon was observed at frequencies lying from the seismic to the ultrasonic range, showing that the role of the slow waves must be taken into account when considering wave propagation in this type of media. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354464]

I. INTRODUCTION

It is well known that variations of reflection coefficients with angle of incidence at the interface between two different media is a useful tool for their characterization. This is very important in the context of exploration geophysics for the study of reservoir rocks saturated by hydrocarbon fluids. There is a vast literature about reflection and transmission coefficients in fluid-saturated porous media (see, Deresiewicz and Rice,[1] Dutta and Odé,[2] Santos *et al.*,[3] Denneman *et al.*,[4] among others). In these works, the medium is assumed to be composed by a porous solid fully saturated by a fluid. The mechanical behavior of the rock matrix, which is generally composed by many different minerals, is represented by a single-phase equivalent elastic or dissipative solid. The same kind of simplification is commonly applied for the description of the pore fluids.

However, there are situations in which it is important to describe independently the behavior of the main components. This requires more detailed formulations involving *multiphase models*. Some important applications are related to the study of seismic and acoustic wave propagation in shaley sandstones, permafrost environments, oceanic or continental sediments containing gas-hydrates, and ultrasonic testing of frozen foods.

A Biot-type three-phase theory describing wave propagation in partially frozen porous media was first derived by Leclaire *et al.*[5] for the case of uniform porosity. The formulation involves two solid phases which are assumed to have no direct contact and one single-phase fluid (water). This model predicts the existence of three compressional waves, referred to as p_I, p_{II}, and p_{III} compressional waves, and two shear waves, referred to as s_I and s_{II}, respectively. Later, Carcione and Tinivella[6] included in that model the interaction between the solid phases and grain cementation with decreasing temperature. More recently, Carcione *et al.*[7] and Santos *et al.*[8] extended and applied that theory to the case of variable porosity.

These models are based on the assumptions of *phase connectivity* for solids and fluids, the validity of the continuous medium description, and *weak coupling* between the solid phases. This means that both solids are nonwelded and consequently they can suffer different displacements and deformations under an applied stress. This assumption also implies that there exists an interchange of strain and kinetic energies between the different constituents.

The computation of the reflection and transmission coefficients of elastic waves in these kinds of media was first presented by Carcione and Tinivella.[6] However, the coefficients were computed using an effective single-phase vis-

---

[a]Electronic mail:grubino@fcaglp.unlp.edu.ar

coelastic model, which ignores the presence of slow waves and neglects the energy conversions among the different wave modes.

This encouraged us to obtain the amplitude and energy reflection and transmission coefficients for waves striking with an arbitrary angle at a plane boundary within an isotropic composite porous saturated solid, using the model developed by Santos *et al.*[8] generalizing the ideas given elsewhere.[2,3] This also led us to define an appropriate energy flux (Umov-Poynting) vector for this model and to analyze its relationship with the energy coefficients.

The application of the procedure is illustrated by considering the case of a plane p_I compressional wave arriving at a plane interface defined by a change in ice content in the pores within a sample of water saturated partially frozen Berea sandstone. The results show that a significant part of the incident energy is partitioned at the interface by mode conversion, generating either reflected or transmitted slow waves. Consequently, the role of these types of waves must be taken into account when considering wave propagation in fluid-saturated composite poroviscoelastic materials, even for frequencies lying in the seismic range.

II. REVIEW OF THE MODEL

Let Q be an elementary cube of porous material composed of two solid phases, referred to by the subscripts or superscripts 1 and 3, saturated by a fluid phase indicated by the subscript or superscript 2, so that $Q = Q_1 \cup Q_2 \cup Q_3$. From now on, such type of material will be referred to as a *composite Biot medium*. Let V_i denote the volume of the phase Q_i and V_b and V_{sm} the bulk volume of Q and the solid matrix $Q_{sm} = Q_1 \cup Q_3$, so that

$$V_{sm} = V_1 + V_3, \quad V_b = V_1 + V_2 + V_3.$$

Let $S_1 = V_1 / V_{sm}$ and $S_3 = V_3 / V_{sm}$ denote the two solid fractions of the composite matrix and define $\phi = V_2 / V_b$, $\phi_1 = V_1 / V_b$, and $\phi_3 = V_3 / V_b$.

Following Santos *et al.*,[8] denoting by ω the angular frequency and $\mathbf{x} = (x, y, z)$ the position of the particle, let $u^{(1)} = u^{(1)}(\mathbf{x}, \omega), \tilde{u}^{(2)} = \tilde{u}^{(2)}(\mathbf{x}, \omega)$, and $u^{(3)} = u^{(3)}(\mathbf{x}, \omega)$ be the time-Fourier transforms of the averaged solid and fluid displacements over the bulk material and the relative fluid flow vector $u^{(2)}$, given by

$$u^{(2)} = u^{(2)}(\mathbf{x}, \omega) = \phi(\tilde{u}^{(2)} - S_1 u^{(1)} - S_3 u^{(3)}), \quad (2.1)$$

with $\zeta = -\nabla \cdot u^{(2)}$ representing the change in fluid content. Also set $u = (u^{(1)}, u^{(2)}, u^{(3)})^T$ and $v = (v^{(1)}, v^{(2)}, v^{(3)})^T$ with $v^{(j)} = i\omega u^{(j)}$ being the corresponding particle velocities.

A. Isotropic stress-strain relations

Next, denote by $\tilde{\sigma}_{jk}^{(1)}$ and $\tilde{\sigma}_{jk}^{(3)}$ the time-Fourier transform of the stress tensors in Q_1 and Q_3 averaged over the bulk material Q, respectively, and let p_f denote the time-Fourier transform of the fluid pressure. Assuming statistical isotropy, let us also introduce the tensors

$$\sigma_{jk}^{(1)} = \tilde{\sigma}_{jk}^{(1)} - S_1 \phi p_f \delta_{jk}, \quad \sigma_{jk}^{(3)} = \tilde{\sigma}_{jk}^{(3)} - S_3 \phi p_f \delta_{jk}, \quad (2.2)$$

associated with the total stresses in Q_1 and Q_3, respectively, taking into account that for the isotropic case the volume fractions S_1, S_3 are equivalent to surface fractions. The linear isotropic stress-strain relations stated in the space-frequency domain are given by[8]

$$\sigma_{jk}^{(1)}(u) = [K_{G1}\theta_1 - B_1 \zeta + B_3 \theta_3]\delta_{jk} + 2\mu_1 d_{jk}^{(1)} + \mu_{13} d_{jk}^{(3)},$$

$$\sigma_{jk}^{(3)}(u) = [K_{G3}\theta_3 - B_2 \zeta + B_3 \theta_1]\delta_{jk} + 2\mu_3 d_{jk}^{(3)} + \mu_{13} d_{jk}^{(1)},$$

$$p_f(u) = -B_1 \theta_1 - B_2 \theta_3 + K_{av}\zeta, \quad (2.3)$$

where $d_{jk}^{(m)} = \epsilon_{jk}(u^{(m)}) - \frac{1}{3}\theta_m \delta_{jk}$, $m = 1, 3$, is the deviatoric strain tensor in Q_{sm}, and $\epsilon_{jk}(u^{(m)})$ denotes linear strain tensor in Q_m, with linear invariant $\theta_m = \epsilon_{jj}(u^{(m)})$.

Santos *et al.*[8] have stated the constitutive relations (2.3) in the space-time domain with real elastic moduli, which were computed in terms of the properties of the individual solid and fluid phases. In this work, some of these moduli will be assumed to be complex and frequency dependent in order to include linear viscoelasticity in the formulation, as described in Appendix A. It must be emphasized that under the weak coupling assumption between the two solids, two different stress tensors are needed to describe the stress states in the solid phases. This constitutes a generalization of the classic Biot's theory, in which the different rock minerals are treated as a single phase solid.

B. The equations of motion

Let the positive definite mass matrix $\mathcal{P} = \mathcal{P}(\omega)$ and the non-negative dissipation matrix $\mathcal{B} = \mathcal{B}(\omega)$ be defined by

$$\mathcal{P} = \begin{bmatrix} p_{11}I & p_{12}I & p_{13}I \\ p_{12}I & p_{22}I & p_{23}I \\ p_{13}I & p_{23}I & p_{33}I \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} f_{11}I & -f_{12}I & -f_{11}I \\ -f_{12}I & f_{22}I & f_{12}I \\ -f_{11}I & f_{12}I & f_{11}I \end{bmatrix},$$

where I denotes the identity matrix in $\mathbb{R}^{3\times3}$. The non-negative coefficients $p_{ij} = p_{ij}(\omega)$, $f_{ij} = f_{ij}(\omega)$, can be computed as explained in Appendix B. Next, let $\mathcal{L}(u)$ be the second-order differential operator defined by

$$\mathcal{L}(u) = (\nabla \cdot \sigma^{(1)}(u), -\nabla p_f(u), \nabla \cdot \sigma^{(3)}(u))^T.$$

Then, the equations of motion of a composite Biot medium, stated in the space-frequency domain, can be written in the form[8]

$$-\omega^2 \mathcal{P}u(\mathbf{x}, \omega) + i\omega \mathcal{B}u(\mathbf{x}, \omega) = \mathcal{L}(u(\mathbf{x}, \omega)), \quad (2.4)$$

which is equivalent to considering an oscillatory time dependence of the form $e^{i\omega t}$. Finally, introducing the matrix $\mathcal{R} = \mathcal{P} - (i/\omega)\mathcal{B}$, we can write

$$-\omega^2 \mathcal{R}u(\mathbf{x}, \omega) = \mathcal{L}(u(\mathbf{x}, \omega)). \quad (2.5)$$

C. Plane wave analysis

Let us consider a monochromatic plane compressional wave of frequency ω in the (x,y) plane. The displacements $u^{(m)}$, $m=1,2,3$, can be written in terms of scalar potentials in the form:

$$u^{(1)} = \nabla \varphi^{(1)} = \nabla(A_P e^{-i\mathbf{K_P} \cdot \mathbf{x}}),$$

$$u^{(2)} = \nabla \varphi^{(2)} = \nabla(B_P e^{-i\mathbf{K_P} \cdot \mathbf{x}}),$$

$$u^{(3)} = \nabla \varphi^{(3)} = \nabla(C_P e^{-i\mathbf{K_P} \cdot \mathbf{x}}), \tag{2.6}$$

where $\mathbf{K_P}=(n_P, l_P)$ is a complex wave vector and $\mathbf{x}=(x,y)$ is the position vector in $\mathbb{R}^2$. Introducing Eq. (2.3) in Eq. (2.4), assuming spatially constant coefficients and using Eq. (2.6) the following 3×3 linear system of equations is obtained:

$$\mathcal{M}^P \cdot c^P = \bar{0}, \tag{2.7}$$

where $c^P = (A_P, B_P, C_P)^T \in \mathbb{C}^{3 \times 1}$ and $\mathcal{M}^P = (\mathcal{M}^P)_{lm} \in \mathbb{C}^{3 \times 3}$ is a symmetric matrix given by

$$(\mathcal{M}^P)_{11} = p_{11} i\omega^2 + f_{11}\omega - K_{G1} i K_P^2 - \tfrac{4}{3}\mu_1 i K_P^2,$$

$$(\mathcal{M}^P)_{12} = p_{12} i\omega^2 - f_{12}\omega - B_1 K_P^2 i,$$

$$(\mathcal{M}^P)_{13} = p_{13}\omega^2 i - f_{11}\omega - B_3 K_P^2 i - \tfrac{2}{3}\mu_{13} i K_P^2,$$

$$(\mathcal{M}^P)_{22} = p_{22}\omega^2 i + f_{22}\omega - K_{av} K_P^2 i,$$

$$(\mathcal{M}^P)_{23} = p_{23}\omega^2 i + f_{12}\omega - B_2 K_P^2 i,$$

$$(\mathcal{M}^P)_{33} = p_{33}\omega^2 i + f_{11}\omega - K_{G3} K_P^2 i - \tfrac{4}{3}\mu_3 i K_P^2, \tag{2.8}$$

with $K_P^2 = n_P^2 + l_P^2$. To obtain nontrivial solutions A_P, B_P, C_P of Eq. (2.7) the determinant of $\mathcal{M}^P$ must be zero, i.e.,

$$\det(\mathcal{M}^P) = P_3(Z) = 0, \tag{2.9}$$

where $Z = K_P^2$ and $P_3(Z)$ is a cubic polynomial in Z. For any given frequency ω, the displacement amplitude of any phase must decay with the distance, which implies that there are only three physically meaningful solutions for the modulus of the wave vector $K_P = \pm\sqrt{Z}$, having negative imaginary parts, corresponding to the p_I, p_{II}, and p_{III} compressional waves,[8] in agreement with Leclaire et al.[5]

For shear (rotational) waves, if $\mathbf{K_S}=(n_S, l_S)$ denotes the corresponding complex wave vector and $\breve{e}_1$, $\breve{e}_2$, $\breve{e}_3$ denote the unit vectors along the x, y, z Cartesian axis, respectively, then

$$u^{(1)} = -\nabla \times \psi^{(1)} = -\nabla \times (A_S e^{-i\mathbf{K_S} \cdot \mathbf{x}} \breve{e}_3),$$

$$u^{(2)} = -\nabla \times \psi^{(2)} = -\nabla \times (B_S e^{-i\mathbf{K_S} \cdot \mathbf{x}} \breve{e}_3),$$

$$u^{(3)} = -\nabla \times \psi^{(3)} = -\nabla \times (C_S e^{-i\mathbf{K_S} \cdot \mathbf{x}} \breve{e}_3), \tag{2.10}$$

and proceeding in a similar fashion, the following linear system is obtained:

$$\mathcal{M}^S \cdot c^S = \bar{0}, \tag{2.11}$$

where $\mathcal{M}^S \in \mathbb{C}^{3 \times 3}$ is a symmetric matrix with entries

$$(\mathcal{M}^S)_{11} = -p_{11} i\omega^2 - f_{11}\omega + \mu_1 i K_S^2,$$

$$(\mathcal{M}^S)_{12} = -p_{12} i\omega^2 + f_{12}\omega,$$

$$(\mathcal{M}^S)_{13} = -p_{13} i\omega^2 + f_{11}\omega + \tfrac{1}{2}\mu_{13} i K_S^2,$$

$$(\mathcal{M}^S)_{22} = -p_{22} i\omega^2 - f_{22}\omega,$$

$$(\mathcal{M}^S)_{23} = -p_{23} i\omega^2 - f_{12}\omega,$$

$$(\mathcal{M}^S)_{33} = -p_{33} i\omega^2 - f_{11}\omega + \mu_3 i K_S^2, \tag{2.12}$$

and $c^S = (A_S, B_S, C_S)^T \in \mathbb{C}^{3 \times 1}$. As before, to obtain nontrivial solutions of Eq. (2.11),

$$P_2(Z) = \det(\mathcal{M}^S) = 0 \tag{2.13}$$

must be solved, with $P_2(Z)$ being a quadratic polynomial in $Z = K_S^2 = n_S^2 + l_S^2$. Equation (2.13) yields only two physically meaningful square roots of Z, with negative imaginary parts, associated with the s_I and s_{II} shear waves.[8]

The p_I and s_I waves correspond to the classical fast longitudinal and transversal waves propagating in elastic or viscoelastic isotropic solids. The additional slow modes are waves strongly attenuated in the low frequency range, related to motions out of phase of the different components. The experimental observation of some of these slow waves was reported by Leclaire et al.[9]

Both for compressional and shear waves, the corresponding phase velocities c_i and attenuation α_i (measured in dB/Hz s) are given by

$$\alpha_i = 2\pi \cdot 8.685\,889 |\text{Im}(K_i)| / |\text{Re}(K_i)|,$$

$$c_i = \omega / |\text{Re}(K_i)|, \quad i = p_I, p_{II}, p_{III}, s_I, s_{II}. \tag{2.14}$$

III. REFLECTION AND TRANSMISSION COEFFICIENTS AT A PLANE INTERFACE

Let us consider an infinitely large plane interface Γ defined by the equation $y=0$ (i.e., normal to $\breve{e}_2$) between two composite Biot half-spaces denoted as Ω_u, Ω_d, where Ω_d is located below Ω_u. A plane p_I compressional wave (of frequency ω) propagating in the plane (x,y) arrives at Γ from Ω_u with an angle θ_{i,p_I} with respect to the normal. Such a perturbation will generate three compressional waves and two shear waves in both Ω_u and Ω_d. The incident, reflected, and transmitted waves will be represented using potentials. For the incident wave p_I, let

$$\varphi_i^1 = A_{i,p_I} e^{-i\mathbf{K_{i,p_I}} \cdot \mathbf{x}},$$

$$\varphi_i^2 = B_{i,p_I} e^{-i\mathbf{K_{i,p_I}} \cdot \mathbf{x}},$$

$$\varphi_i^3 = C_{i,p_I} e^{-i\mathbf{K_{i,p_I}} \cdot \mathbf{x}} \tag{3.1}$$

be the scalar potentials for the solid 1, the fluid, and the solid 3, respectively. Also, let φ_q^j and ψ_q^j be the compressional and shear potentials for the solids 1 and 3 ($j=1$ and $j=3$, respectively) and for the fluid displacement ($j=2$), with $q=r$ for the

reflected waves and $q=t$ for the transmitted waves. They are given by

$$\varphi_q^1 = A_{q,p_I}e^{-i\mathbf{K_{q,p_I} \cdot x}} + A_{q,p_{II}}e^{-i\mathbf{K_{q,p_{II}} \cdot x}} + A_{q,p_{III}}e^{-i\mathbf{K_{q,p_{III}} \cdot x}},$$

$$\psi_q^1 = [A_{q,s_I}e^{-i\mathbf{K_{q,s_I} \cdot x}} + A_{q,s_{II}}e^{-i\mathbf{K_{q,s_{II}} \cdot x}}]\breve{e}_3,$$

$$\varphi_q^2 = B_{q,p_I}e^{-i\mathbf{K_{q,p_I} \cdot x}} + B_{q,p_{II}}e^{-i\mathbf{K_{q,p_{II}} \cdot x}} + B_{q,p_{III}}e^{-i\mathbf{K_{q,p_{III}} \cdot x}},$$

$$\psi_q^2 = [B_{q,s_I}e^{-i\mathbf{K_{q,s_I} \cdot x}} + B_{q,s_{II}}e^{-i\mathbf{K_{q,s_{II}} \cdot x}}]\breve{e}_3,$$

$$\varphi_q^3 = C_{q,p_I}e^{-i\mathbf{K_{q,p_I} \cdot x}} + C_{q,p_{II}}e^{-i\mathbf{K_{q,p_{II}} \cdot x}} + C_{q,p_{III}}e^{-i\mathbf{K_{q,p_{III}} \cdot x}},$$

$$\psi_q^3 = [C_{q,s_I}e^{-i\mathbf{K_{q,s_I} \cdot x}} + C_{q,s_{II}}e^{-i\mathbf{K_{q,s_{II}} \cdot x}}]\breve{e}_3. \quad (3.2)$$

In Eqs. (3.1) and (3.2),

$$\mathbf{K_{q,j}} = (n_{q,j}, l_{q,j}) = K_{q,j}(\sin\theta_{q,j}, \cos\theta_{q,j}),$$

$$q = i, r, t, \quad j = p_I, p_{II}, p_{III}, s_I, s_{II},$$

are the complex wave vectors for each type of wave, with $K_{q,j}$ satisfying equations of the form (2.9) or (2.13) in Ω_u for $q=i,r$ and in Ω_d for $q=t$. Also, $\theta_{q,j}$ denotes the incidence, reflection, or transmission angle for the corresponding wave.

The displacement vectors in the three phases in Ω_u and Ω_d, denoted $u^{(m,u)}, u^{(m,d)}$, are given by

$$u^{(m,u)} = (u_x^{(m,u)}, u_y^{(m,u)}) = \nabla\varphi_i^m + \nabla\varphi_r^m - \nabla \times \psi_r^m$$

$$= u_{(i,p_I)}^{(m)} + u_{(r,p_I)}^{(m)} + u_{(r,p_{II})}^{(m)} + u_{(r,p_{III})}^{(m)} + u_{(r,s_I)}^{(m)} + u_{(r,s_{II})}^{(m)},$$

$$u^{(m,d)} = (u_x^{(m,d)}, u_y^{(m,d)}) = \nabla\varphi_t^m - \nabla \times \psi_t^m$$

$$= u_{(t,p_I)}^{(m)} + u_{(t,p_{II})}^{(m)} + u_{(t,p_{III})}^{(m)} + u_{(t,s_I)}^{(m)} + u_{(t,s_{II})}^{(m)}, \quad m = 1, 2, 3. \quad (3.3)$$

Here, $u_{(q,j)}^{(m,u)}, u_{(q,j)}^{(m,d)}$ $(q=i,r,t, j=p_I, p_{II}, p_{III}, s_I, s_{II})$ denote the parts of the displacement vectors $(u^{(m,u)}$ and $u^{(m,d)})$ associated with the incident, reflected, or transmitted j wave.

The natural boundary conditions at the interface Γ (i.e., at $y=0$) are the continuity of displacements and generalized forces, as an extension of those given by Dutta and Odé:[2]

$$u_x^{(1,u)} = u_x^{(1,d)}, \quad (3.4a)$$

$$u_y^{(1,u)} = u_y^{(1,d)}, \quad (3.4b)$$

$$u_x^{(3,u)} = u_x^{(3,d)}, \quad (3.4c)$$

$$u_y^{(3,u)} = u_y^{(3,d)}, \quad (3.4d)$$

$$u_y^{(2,u)} = u_y^{(2,d)}, \quad (3.4e)$$

$$p_f^{(u)} = p_f^{(d)}, \quad (3.4f)$$

$$\sigma_{21}^{(1,u)} = \sigma_{21}^{(1,d)}, \quad (3.4g)$$

$$\sigma_{22}^{(1,u)} = \sigma_{22}^{(1,d)}, \quad (3.4h)$$

$$\sigma_{21}^{(3,u)} = \sigma_{21}^{(3,d)}, \quad (3.4i)$$

$$\sigma_{22}^{(3,u)} = \sigma_{22}^{(3,d)}, \quad (3.4j)$$

where, from now on, the superscripts (u) and (d) denote, respectively, the values of any variable in Ω_u and Ω_d.

The reflection and transmission coefficients R_j and T_j, $j=p_I, p_{II}, p_{III}, s_I, s_{II}$, can be defined for each wave as the ratio of the corresponding displacement amplitude of the solid 1 and that of the incident wave, i.e.,

$$R_j = A_{r,j}K_{r,j}/A_{i,p_I}K_{i,p_I},$$

$$T_j = A_{t,j}K_{t,j}/A_{i,p_I}K_{i,p_I}. \quad (3.5)$$

Since the amplitude A_{i,p_I} as well as the incidence angle θ_{i,p_I}, or equivalently the incident complex wave vector, are assumed to be known, to compute the reflection and transmission coefficients (3.5) the amplitudes $A_{r,j}, A_{t,j}$ must be obtained. For this purpose, first note that each triplet $(u_{(q,j)}^{(1)}, u_{(q,j)}^{(2)}, u_{(q,j)}^{(3)})$ in Eq. (3.3) must individually satisfy either Eq. (2.7) or Eq. (2.11). Thus, taking into account that $\mathcal{M}^{\mathcal{P}}$ and $\mathcal{M}^{\mathcal{S}}$ are singular matrices, the amplitudes of the different types of waves in the fluid phase and solid 3 phase can be written in terms of that of the solid 1 phase by means of the following relations:

$$\gamma_i^B = \frac{B_{i,p_I}}{A_{i,p_I}}, \quad \gamma_i^C = \frac{C_{i,p_I}}{A_{i,p_I}},$$

$$\gamma_{q,j}^B = \frac{B_{q,j}}{A_{q,j}}, \quad \gamma_{q,j}^C = \frac{C_{q,j}}{A_{q,j}}, \quad q = r, t, \quad j = p_I, p_{II}, p_{III}, s_I, s_{II}. \quad (3.6)$$

Next, using the boundary conditions (3.4) and the relations (3.6) the desired amplitudes $A_{q,j}$ are computed. First, from Eqs. (3.3) and (3.4a), it is easy to see that

$$n_{i,p_I} = n_{r,p_I} = n_{r,p_{II}} = n_{r,p_{III}} = n_{r,s_I} = n_{r,s_{II}} = n_{t,p_I} = n_{t,p_{II}}$$

$$= n_{t,p_{III}} = n_{t,s_I} = n_{t,s_{II}}, \quad (3.7)$$

which is a generalization of Snell's law for this model, and it allows us to obtain the reflection and transmission angles $\theta_{q,j}$ for each type of wave as functions of the incidence angle.

Now, using Eqs. (3.3), (3.4), (3.6), and (3.7), the following 10×10 linear system of equations for the unknowns $A_{m,j}$ is obtained, with right-hand side Y_m, $m=1, \ldots, 10$:

$$-\sum_{j=p_I, p_{II}, p_{III}} n_{i,p_I}(A_{r,j} - A_{t,j}) + \sum_{\alpha=s_I, s_{II}} (l_{r,\alpha}A_{r,\alpha} - l_{t,\alpha}A_{t,\alpha})$$

$$= Y_1,$$

$$\sum_{j=p_I, p_{II}, p_{III}} (-l_{r,j}A_{r,j} + l_{t,j}A_{t,j}) - \sum_{\alpha=s_I, s_{II}} n_{i,p_I}(A_{r,\alpha} - A_{t,\alpha}) = Y_2,$$

$$\sum_{j=p_I, p_{II}, p_{III}} n_{i,p_I}(-\gamma_{r,j}^C A_{r,j} + \gamma_{t,j}^C A_{t,j}) + \sum_{\alpha=s_I, s_{II}} (\gamma_{r,\alpha}^C l_{r\alpha}A_{r,\alpha}$$

$$- \gamma_{t,\alpha}^C l_{t,\alpha}A_{t,\alpha}) = Y_3,$$

$$\sum_{j=p_I,p_{II},p_{III}} (-\gamma_{r,j}^C l_{r,j} A_{r,j} + \gamma_{t,j}^C l_{t,j} A_{t,j}) + \sum_{\alpha=s_I,s_{II}} n_{i,p_I}(-\gamma_{r,\alpha}^C A_{r,\alpha}$$
$$+ \gamma_{t,\alpha}^C A_{t,\alpha}) = Y_4,$$

$$\sum_{j=p_I,p_{II},p_{III}} (-\gamma_{r,j}^B l_{r,j} A_{r,j} + \gamma_{t,j}^B l_{t,j} A_{t,j}) + \sum_{\alpha=s_I,s_{II}} n_{i,p_I}(-\gamma_{r,\alpha}^B A_{r,\alpha}$$
$$+ \gamma_{t,\alpha}^B A_{t,\alpha}) = Y_5,$$

$$\sum_{j=p_I,p_{II},p_{III}} [K_{r,j}^2[B_1^u + \gamma_{r,j}^C B_2^u + K_{av}^u \gamma_{r,j}^B]A_{r,j} - K_{t,j}^2[B_1^d + \gamma_{t,j}^C B_2^d$$
$$+ K_{av}^d \gamma_{t,j}^B]A_{t,j}] = Y_6,$$

$$\sum_{j=p_I,p_{II},p_{III}} n_{i,p_I}[l_{r,j}[-2\mu_1^u - \mu_{13}^u \gamma_{r,j}^C]A_{r,j} + l_{t,j}[2\mu_1^d$$
$$+ \mu_{13}^d \gamma_{t,j}^C]A_{t,j}]$$

$$+ \sum_{\alpha=s_I,s_{II}} \left[(l_{r,\alpha}^2 - n_{i,p_I}^2)\left[\mu_1^u + \frac{1}{2}\mu_{13}^u \gamma_{r,\alpha}^C\right]A_{r,\alpha} + (n_{i,p_I}^2 - l_{t,\alpha}^2)\right.$$
$$\left. \times \left[\mu_1^d + \frac{1}{2}\mu_{13}^d \gamma_{t,\alpha}^C\right]A_{t,\alpha}\right] = Y_7,$$

$$\sum_{j=p_I,p_{II},p_{III}} \left[\left[K_{r,j}^2\left(\frac{2}{3}\mu_1^u - K_{G1}^u - B_1^u \gamma_{r,j}^B + \gamma_{r,j}^C\left(\frac{1}{3}\mu_{13}^u - B_3^u\right)\right)\right.\right.$$
$$\left. - l_{r,j}^2\left(2\mu_1^u + \mu_{13}^u \gamma_{r,j}^C\right)\right]A_{r,j}$$

$$+ \left[K_{t,j}^2\left(K_{G1}^d - \frac{2}{3}\mu_1^d + B_1^d \gamma_{t,j}^B + \gamma_{t,j}^C\left(B_3^d - \frac{1}{3}\mu_{13}^d\right)\right)\right.$$
$$\left.\left. + l_{t,j}^2\left(2\mu_1^d + \mu_{13}^d \gamma_{t,j}^C\right)\right]A_{t,j}\right]$$

$$+ \sum_{\alpha=s_I,s_{II}} [n_{i,p_I}l_{r,\alpha}[-2\mu_1^u - \mu_{13}^u \gamma_{r,\alpha}^C]A_{r,\alpha} + n_{i,p_I}l_{t,\alpha}[2\mu_1^d$$
$$+ \mu_{13}^d \gamma_{t,\alpha}^C]A_{t,\alpha}] = Y_8,$$

$$\sum_{j=p_I,p_{II},p_{III}} n_{i,p_I}[l_{r,j}(-2\mu_3^u \gamma_{r,j}^C - \mu_{13}^u)A_{r,j} + l_{t,j}(2\mu_3^d \gamma_{t,j}^C$$
$$+ \mu_{13}^d)A_{t,j}]$$

$$+ \sum_{\alpha=s_I,s_{II}} \left[(l_{r,\alpha}^2 - n_{i,p_I}^2)\left(\mu_3^u \gamma_{r,\alpha}^C + \frac{1}{2}\mu_{13}^u\right)A_{r,\alpha} - (l_{t,\alpha}^2 - n_{i,p_I}^2)\right.$$
$$\left. \times \left(\mu_3^d \gamma_{t,\alpha}^C + \frac{1}{2}\mu_{13}^d\right)A_{t,\alpha}\right] = Y_9,$$

$$\sum_{j=p_I,p_{II},p_{III}} \left[\left[K_{r,j}^2\left(\frac{1}{3}\mu_{13}^u - B_3^u - B_2^u \gamma_{r,j}^B + \gamma_{r,j}^C\left(\frac{2}{3}\mu_3^u - K_{G3}^u\right)\right)\right.\right.$$
$$\left. - l_{r,j}^2(\mu_{13}^u + 2\mu_3^u \gamma_{r,j}^C)\right]A_{r,j}$$

$$+ \left[K_{t,j}^2\left(B_3^d - \frac{1}{3}\mu_{13}^d + B_2^d \gamma_{t,j}^B + \left(K_{G3}^d - \frac{2}{3}\mu_3^d\right)\gamma_{t,j}^C\right)\right.$$
$$\left.\left. + l_{t,j}^2\left(\mu_{13}^d + 2\mu_3^d \gamma_{t,j}^C\right)\right]A_{t,j}\right]$$

$$+ \sum_{\alpha=s_I,s_{II}} [n_{i,p_I}l_{r,\alpha}(-2\mu_3^u \gamma_{r,\alpha}^C - \mu_{13}^u)A_{r,\alpha} + n_{i,p_I}l_{t,\alpha}(2\mu_3^d \gamma_{t,\alpha}^C$$
$$+ \mu_{13}^d)A_{t,\alpha}] = Y_{10}, \quad (3.8)$$

where using Snell's law (3.7)

$$l_{r,j} = (K_{r,j}^2 - n_{i,p_I}^2)^{1/2}, \quad l_{t,j} = -(K_{t,j}^2 - n_{i,p_I}^2)^{1/2}, \quad (3.9)$$

$$j = p_I, p_{II}, p_{III}, s_I, s_{II},$$

and

$$Y_1 = n_{i,p_I}A_{i,p_I}, \quad Y_2 = l_{i,p_I}A_{i,p_I}, \quad Y_3 = \gamma_i^C n_{i,p_I}A_{i,p_I},$$

$$Y_4 = \gamma_i^C l_{i,p_I}A_{i,p_I},$$

$$Y_5 = \gamma_i^B A_{i,p_I}l_{i,p_I}, \quad Y_6 = A_{i,p_I}K_{i,p_I}^2[-B_1^u - B_2^u \gamma_i^C - K_{av}^u \gamma_i^B],$$

$$Y_7 = (2\mu_1^u + \mu_{13}^u \gamma_i^C)n_{i,p_I}l_{i,p_I}A_{i,p_I},$$

$$Y_8 = A_{i,p_I}\left[K_{i,p_I}^2\left(K_{G1}^u - \frac{2}{3}\mu_1^u + B_1^u \gamma_i^B + \left(B_3^u - \frac{1}{3}\mu_{13}^u\right)\gamma_i^C\right)\right.$$
$$\left. + l_{i,p_I}^2(2\mu_1^u + \mu_{13}^u \gamma_i^C)\right],$$

$$Y_9 = n_{i,p_I}l_{i,p_I}(2\mu_3^u \gamma_i^C + \mu_{13}^u)A_{i,p_I},$$

$$Y_{10} = A_{i,p_I}\left[K_{i,p_I}^2\left(B_3^u - \frac{1}{3}\mu_{13}^u + B_2^u \gamma_i^B + \gamma_i^C\left(K_{G3}^u - \frac{2}{3}\mu_3^u\right)\right)\right.$$
$$\left. + l_{i,p_I}^2(\mu_{13}^u + 2\mu_3^u \gamma_i^C)\right].$$

The signs in Eq. (3.9) are chosen according to the propagating direction (upgoing or downgoing) of the different waves. Once the linear system (3.8) is solved, the amplitude reflection and transmission coefficients R_j and T_j at the interface Γ can be computed from Eq. (3.5).

IV. ENERGY CONSIDERATIONS IN THE COMPOSITE MEDIUM

It is also useful and physically important to compute the reflection and transmission coefficients in terms of the energy flux across the interface. Since our formulation is stated in the space-frequency domain, and the model coefficients are frequency dependent, it is necessary to derive an expression for the *energy flux vector* for the composite medium in the frequency domain.

A. Definition of the Umov-Poynting vector

Following the usual procedure, take the dot product of Eq. (2.5) with v^{*T} (where the superscript * denotes the complex conjugate) and integrate over a volume V. Using integration by parts in the $\mathcal{L}(u)$ term and the symmetry properties of the tensors $\sigma_{ij}^{(1)}(u), \sigma_{ij}^{(3)}(u)$, the following equation is obtained:

$$i\omega \int_V v^{*T}\mathcal{R}v\,dV - \int_V \left(\frac{\partial \sigma_{ij}^{(1)}(u)}{\partial x_i}v_j^{(1)*} \right.$$

$$\left. - \frac{\partial p_f(u)}{\partial x_i}v_i^{(2)*} + \frac{\partial \sigma_{ij}^{(3)}(u)}{\partial x_i}v_j^{(3)*} \right)dV$$

$$= i\omega \int_V v^{*T}\mathcal{R}v\,dV + \int_V (\sigma_{ij}^{(1)}(u)\epsilon_{ij}(v^{(1)*}))$$

$$- p_f(u)\,\nabla \cdot v^{(2)*} + \sigma_{ij}^{(3)}(u)\epsilon_{ij}(v^{(3)*}))dV$$

$$- \int_{\partial V} [(\sigma_{ij}^{(1)}(u)v_j^{(1)*} - p_f(u)v_i^{(2)*} + \sigma_{ij}^{(3)}(u)v_j^{(3)*})\nu_i]dS$$

$$= 0, \tag{4.1}$$

where $\nu=(\nu_i)$ denotes the unit outer normal to ∂V.

Let $\mathcal{S}\in\mathbb{C}^{13\times13}$ be the complex symmetric frequency-dependent stiffness matrix, having as entries the viscoelastic moduli of the medium. Then the stress-strain relations (2.3) can be written in the form

$$\sigma = \mathcal{S}\epsilon, \tag{4.2}$$

where $\sigma=[\sigma_{ij}^{(1)}(u),\sigma_{ij}^{(3)}(u),p_f(u)]^T\in\mathbb{C}^{13\times1}$ and $\epsilon=[\epsilon_{ij}(u^{(1)}),\epsilon_{ij}(u^{(3)}),\zeta]^T\in\mathbb{C}^{13\times1}$, $1\le i\le j\le3$, are the stress and strain arrays, respectively.

Next, use Eq. (4.2) in Eq. (4.1), to derive the relation

$$i\omega \int_V v^{*T}\mathcal{R}v\,dV - i\omega \int_V \epsilon^{*T}\mathcal{S}\epsilon\,dV = \int_{\partial V} [(\sigma_{ij}^{(1)}(u)v_j^{(1)*}$$

$$- p_f(u)v_i^{(2)*} + \sigma_{ij}^{(3)}(u)v_j^{(3)*})\nu_i]dS. \tag{4.3}$$

Splitting the terms in Eq. (4.3) into their real and imaginary parts leads to

$$i\omega \int_V 2(\mathcal{W} - \mathcal{T})dV - \int_V (\hat{\mathcal{D}}_\mathcal{T} + \hat{\mathcal{D}}_\mathcal{W})dV = \int_{\partial V} \mathbf{P}\cdot\nu\,dS$$

$$= \int_V \nabla\cdot\mathbf{P}\,dV, \tag{4.4}$$

where

$$\mathcal{T} = \frac{1}{4}\,\text{Re}(v^{*T}\mathcal{R}v), \quad \mathcal{W} = \frac{1}{4}\,\text{Re}(\epsilon^{*T}\mathcal{S}\epsilon),$$

$$\hat{\mathcal{D}}_\mathcal{T} = -\frac{\omega}{2}\,\text{Im}(v^{*T}\mathcal{R}v), \quad \hat{\mathcal{D}}_\mathcal{W} = \frac{\omega}{2}\,\text{Im}(\epsilon^{*T}\mathcal{S}\epsilon). \tag{4.5}$$

For any complex scalar or tensor A, $\text{Re}(A)$ and $\text{Im}(A)$ denote the real and imaginary parts of A, respectively. The vector $\mathbf{P}=P_k\breve{e}_k$, with components

$$P_k(u,v) = -\frac{1}{2}(\sigma_{kj}^{(1)}(u)v_j^{(1)*} - p_f(u)v_k^{(2)*} + \sigma_{kj}^{(3)}(u)v_j^{(3)*}), \tag{4.6}$$

is the *complex Umov-Poynting vector* for this model, analogous to that given by Carcione[10] for poroviscoelastic media, and can be considered as its formal generalization.

To understand the physical significance of the different terms in Eqs. (4.4) and (4.5), recall that if A is a complex time-harmonic vector having entries of the form $a = a_0 e^{i(\omega t-\beta)}$, and D is a complex symmetric matrix, then, denoting by

$$\langle a \rangle = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} a(t)dt$$

the time average over a cycle, the following relations can be verified:[10]

$$\langle \text{Re}(A^T)\text{Re}(D)\text{Re}(A) \rangle = \frac{1}{2}\,\text{Re}(A^T D A^*),$$

$$\langle \text{Re}(A^T)\text{Im}(D)\text{Re}(A) \rangle = \frac{1}{2}\,\text{Im}(A^T D A^*). \tag{4.7}$$

Since by hypothesis all variables on the right-hand side of Eq. (4.5) are periodic with temporal dependency of the form $e^{i\omega t}$, it is straightforward to see that

$$\mathcal{T} = \frac{1}{2}\langle \text{Re}(v^T)\text{Re}(\mathcal{R})\text{Re}(v) \rangle, \tag{4.8a}$$

$$\mathcal{W} = \frac{1}{2}\langle \text{Re}(\epsilon^T)\text{Re}(\mathcal{S})\text{Re}(\epsilon) \rangle, \tag{4.8b}$$

$$\hat{\mathcal{D}}_\mathcal{T} = -\omega\langle \text{Re}(v^T)\text{Im}(\mathcal{R})\text{Re}(v) \rangle, \tag{4.8c}$$

$$\hat{\mathcal{D}}_\mathcal{W} = \omega\langle \text{Re}(\epsilon^T)\text{Im}(\mathcal{S})\,\text{Re}(\epsilon) \rangle. \tag{4.8d}$$

Taking into account that (4.8a) and (4.8c) can be written as

$$\mathcal{T} = \frac{1}{2}\langle \text{Re}(v^T)\mathcal{P}\,\text{Re}(v) \rangle, \quad \hat{\mathcal{D}}_\mathcal{T} = \langle \text{Re}(v^T)\mathcal{B}\,\text{Re}(v) \rangle,$$

it can be concluded that $\mathcal{T}$ is the kinetic energy density and $\hat{\mathcal{D}}_\mathcal{T}$ is the rate of dissipated kinetic energy density. With regards to the stress-strain terms, note that (4.8b) is analogous to the strain energy density for this model,[8] while the term (4.8d) represents the rate of dissipated strain energy density. Thus, Eq. (4.4) represents an energy-balance equation relating the complex Umov-Poynting vector with the energy densities $\mathcal{T}$, $\mathcal{W}$ and their rates of dissipation $\hat{\mathcal{D}}_\mathcal{T}$ and $\hat{\mathcal{D}}_\mathcal{W}$.

Also using Eq. (4.7), the real Umov-Poynting vector $\mathbf{P_R}=P_{Rk}\breve{e}_k$, defined by

$$P_{Rk} = -(\text{Re}(\sigma_{kj}^{(1)}(u))\text{Re}(v_j^{(1)}) - \text{Re}(p_f(u))\text{Re}(v_k^{(2)})$$

$$+ \text{Re}(\sigma_{kj}^{(3)}(u))\text{Re}(v_j^{(3)})) \tag{4.9}$$

satisfies $\langle \mathbf{P_R} \rangle = \text{Re}(\mathbf{P})$, in agreement with other works,[10,11] and this time-average represents the magnitude and direction of the time-averaged power flow. As a consequence of the boundary conditions (3.4), the normal component of the real Umov-Poynting vector (4.9) is continuous at the interface.

Note that the analysis performed in this section (Sec. IV A) was not restricted to the particular case of plane wave propagation.

B. Energy reflection and transmission coefficients

Using the previous results, let us go back to the reflection-transmission problem at a plane interface. The time-average of the normal component of the energy flux, either in Ω_u or in Ω_d, is given by

$$F = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \mathbf{P_R} \cdot \breve{e}_2 dt = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} P_{R2}(u,v) dt. \quad (4.10)$$

Applying the superposition principle stated in Eq. (3.3) to the particle velocities $v^{(m,u)}, v^{(m,d)}$, the generalized stress components and the fluid pressure, F can be splitted into different components associated with the different wave modes. In this sense, following Dutta and Odé[2] (except for a sign), the *partial orthodox fluxes* $F_{k,k}$ are defined as

$$F_{k,k} = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} (-1)(\mathrm{Re}(\sigma_{2j,k}^{(1)}(u))\mathrm{Re}(v_{j,k}^{(1)})$$
$$+ \mathrm{Re}(\sigma_{2j,k}^{(3)}(u))\mathrm{Re}(v_{j,k}^{(3)}) - \mathrm{Re}(p_{f,k}(u))\mathrm{Re}(v_{2,k}^{(2)}))dt,$$
$$j = 1,2, \quad (4.11)$$

and the *interference fluxes* $F_{k,q}$ are given by

$$F_{k,q} = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} (-1)(\mathrm{Re}(\sigma_{2j,k}^{(1)}(u))\mathrm{Re}(v_{j,q}^{(1)})$$
$$+ \mathrm{Re}(\sigma_{2j,k}^{(3)}(u))\mathrm{Re}(v_{j,q}^{(3)}) - \mathrm{Re}(p_{f,k}(u))\mathrm{Re}(v_{2,q}^{(2)})$$
$$+ \mathrm{Re}(\sigma_{2j,q}^{(1)}(u))\mathrm{Re}(v_{j,k}^{(1)}) + \mathrm{Re}(\sigma_{2j,q}^{(3)}(u))\mathrm{Re}(v_{j,k}^{(3)})$$
$$- \mathrm{Re}(p_{f,q}(u))\mathrm{Re}(v_{2,k}^{(2)}))dt, \quad j = 1,2, \quad (4.12)$$

where $k,q = ip_\mathrm{I}, rp_\mathrm{I}, rp_\mathrm{II}, rp_\mathrm{III}, rs_\mathrm{I}, rs_\mathrm{II}$ in Ω_u and $k,q = tp_\mathrm{I}, tp_\mathrm{II}, tp_\mathrm{III}, ts_\mathrm{I}, ts_\mathrm{II}$ in Ω_d denotes the wave associated with the variable and the sum convention is applied on the index j.

In this way, the energy reflection and transmission coefficients are defined as follows:

$$ER_j = \frac{|F_{rj,rj}|}{|F_{ip_\mathrm{I}}|},$$

$$ET_j = \frac{|F_{tj,tj}|}{|F_{ip_\mathrm{I}}|}, \quad j = p_\mathrm{I}, p_\mathrm{II}, p_\mathrm{III}, s_\mathrm{I}, s_\mathrm{II}, \quad (4.13)$$

where

$$F_{ip_\mathrm{I}} = F_{ip_\mathrm{I},ip_\mathrm{I}} + F_{ip_\mathrm{I},rp_\mathrm{I}} + F_{ip_\mathrm{I},rp_\mathrm{II}} + F_{ip_\mathrm{I},rp_\mathrm{III}} + F_{ip_\mathrm{I},rs_\mathrm{I}} + F_{i,rs_\mathrm{II}}. \quad (4.14)$$

The total reflected and transmitted energy fluxes F_r and F_t are defined as

$$F_j = F_{jp_\mathrm{I},jp_\mathrm{I}} + F_{jp_\mathrm{II},jp_\mathrm{II}} + F_{jp_\mathrm{III},jp_\mathrm{III}} + F_{js_\mathrm{I},js_\mathrm{I}} + F_{js_\mathrm{II},js_\mathrm{II}}$$
$$+ F_{jp_\mathrm{I},jp_\mathrm{II}} + F_{jp_\mathrm{I},jp_\mathrm{III}} + F_{jp_\mathrm{I},js_\mathrm{I}} + F_{jp_\mathrm{I},js_\mathrm{II}} + F_{jp_\mathrm{II},jp_\mathrm{III}}$$
$$+ F_{jp_\mathrm{II},js_\mathrm{I}} + F_{jp_\mathrm{II},js_\mathrm{II}} + F_{jp_\mathrm{III},js_\mathrm{I}} + F_{jp_\mathrm{III},js_\mathrm{II}} + F_{js_\mathrm{I},js_\mathrm{II}}, \quad (4.15)$$

where $j = r, t$. Thus, the continuity of the normal component of the real Umov-Poynting vector implies the energy balance at the interface, which can be stated as

$$F_{ip_\mathrm{I}} + F_r = F_t. \quad (4.16)$$

TABLE I. Material properties of the frozen sandstone model.

Solid grain	Bulk modulus, K_{s1}	38.7 GPa
	Shear modulus, μ_{s1}	39.6 GPa
	Density, ρ_1	2650 kg/m³
	Permeability $\kappa_{1,0}$	1.07×10^{-13} m²
Ice	Bulk modulus, K_{s3}	8.58 GPa
	Shear modulus, μ_{s3}	3.32 GPa
	Density, ρ_3	920 kg/m³
	Permeability $\kappa_{3,0}$	5×10^{-4} m²
Fluid	Bulk modulus, K_f	2.25 GPa
	Density, ρ_2	1000 kg/m³
	Viscosity, η	1 cP
Air	Bulk modulus, K_a	1.5×10^{-4} GPa
	Shear modulus, μ_a	0 GPa

V. NUMERICAL EXAMPLE

Here the formulas derived in Sec. IV are used to compute the energy reflection and transmission coefficients for an incident plane p_I compressional wave striking a plane interface defined by a change in ice content in the pores within a sample of water saturated partially frozen Berea sandstone. The material properties of the system are given in Table I.[6,7] The absolute porosity of the sandstone, defined as the ratio of the volume of the interconnected pores of the rock (V_p) and the total volume of the sample, i.e., $\phi_a = V_p/V_b$, is taken to be equal to 0.18. The bulk water content ϕ was chosen such that the ice content in the pores $\phi_i = V_3/V_p = \phi_3/(1-\phi_1)$ varies from 0.1 in the upper layer to 0.9 in the lower layer.

The calculation of the elastic coefficients requires values for the bulk and shear moduli of the two solid (dry) frames, denoted by $K_{s1,m}, K_{s3,m}, \mu_{s1,m}$, and $\mu_{s3,m}$, respectively (see Appendix A 1). It is assumed that $K_{s1,m} = 14.4$ GPa and that $\mu_{s1,m}$, $\mu_{s3,m}$, and $K_{s3,m}$ can be obtained using a percolation-type model in the form[5,6,8]

$$\mu_{sj,m} = [\mu_{sj,m}^{(\mathrm{max})} - \mu_{sj,m}^0]\left[\frac{\phi_3}{1-\phi_1}\right]^{3.8} + \mu_{sj,m}^0, \quad j = 1,3,$$

$$K_{s3,m} = [K_{s3,m}^{(\mathrm{max})} - K_{s3,m}^0]\left[\frac{\phi_3}{1-\phi_1}\right]^{3.8} + K_{s3,m}^0, \quad (5.1)$$

where $\mu_{s1,m}^{(\mathrm{max})}$, $\mu_{s3,m}^{(\mathrm{max})}$, and $K_{s3,m}^{(\mathrm{max})}$ are computed employing a Kuster and Toksöz model,[8,12] using known values of $K_{s1}, \mu_{s1}, K_{s3}, \mu_{s3}$ for the background medium with inclusions of air, with properties K_a, μ_a (see Table I). The moduli $\mu_{s1,m}^0$, $\mu_{s3,m}^0$, and $K_{s3,m}^0$ are reference values taken to be

$$\mu_{s1,m}^0 = 13.1 \text{ GPa}, \quad K_{s3,m}^0 = \mu_{s3,m}^0 = 0. \quad (5.2)$$

The viscoelastic parameters describing the dissipative behavior of the saturated sandstone are (see Appendix A 2) $T_{1,M} = (2\pi 10^{-5})^{-1}$ ms, $T_{2,M} = (2\pi 10^6)^{-1}$ ms, for $M = K_{G1}, K_{G3}, \mu_1, \mu_3$. For the mean quality factors, we take $\hat{Q}_{K_{G1}} = \hat{Q}_{\mu_1} = 100$ and $\hat{Q}_{K_{G3}} = \hat{Q}_{\mu_3} = 80$. The value of the Kozeny-Carman constant was taken to be 5.[13] For these rock and fluid properties, the phase velocities for the different wave modes given by Eq. (2.14) are presented in Table II for

TABLE II. Phase velocities for the different wave modes (km/S).

	Upper layer ϕ_i=0.1, Freq=20 Hz	Lower layer ϕ_i=0.9, Freq=20 Hz	Upper layer ϕ_i=0.1, Freq=1 MHz	Lower layer ϕ_i=0.9, Freq=1 MHz
$c_{p_{\mathrm{I}}}$	4.038	4.988	4.205	5.172
$c_{p_{\mathrm{II}}}$	0.319	1.727	0.758	1.970
$c_{p_{\mathrm{III}}}$	0.006	0.0007	0.148	0.093
$c_{s_{\mathrm{I}}}$	2.423	3.233	2.54	3.365
$c_{s_{\mathrm{II}}}$	0.039	0.827	0.042	0.927

20 Hz and 1 MHz, corresponding to frequencies used in seismic exploration and laboratory testing, respectively.

Figures 1–4 show the magnitude of the energy coefficients computed using Eq. (4.13) as functions of the incidence angle. Since the energy coefficients related to the p_{III} compressional waves are negligible, for brevity the corresponding curves are not presented. Due to the dissipative behavior of the medium and the inhomogeneous nature of the different wave modes, the analysis of the existence of critical angles requires a rigorous definition. For each transmitted wave mode the critical angle(s) is computed as the angle(s) of incidence beyond which the *corresponding* time-averaged energy flux vector (i.e., the time-average of the real Umov-Poynting vector $\mathbf{P_R}$) becomes parallel to the interface.[10,14] In our case, neglecting the interference fluxes,

this is approximately equivalent to the condition that the orthodox flux $F_{tj,tj}$ vanishes (i.e., ET_j=0). Figure 3(a) shows the presence of a critical angle associated to the p_{I} wave at about 55° for both frequencies. This value is very close to the approximate value 54° obtained from the classic Snell's law for two elastic media using the fast velocities $c_{p_{\mathrm{I}}}$ in Table II. As expected, beyond the critical angle, Figs. 1(a) and 3(a) show an important increase in the reflected energy of the p_{I} wave, with a consequent decrease in the transmitted energy for the same kind of wave.

The other wave modes have peaks near the critical angle, which is particularly interesting in the case of the transmission of the p_{II} compressional wave in Fig. 3(b), showing that near this angle of incidence almost a fifth of the incident energy is being converted into this slow wave. As
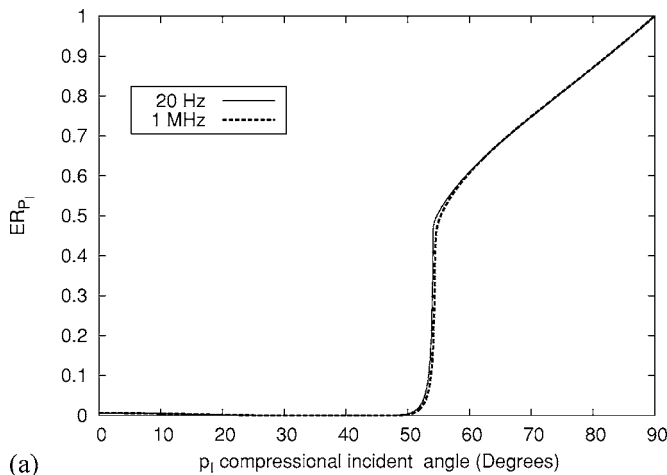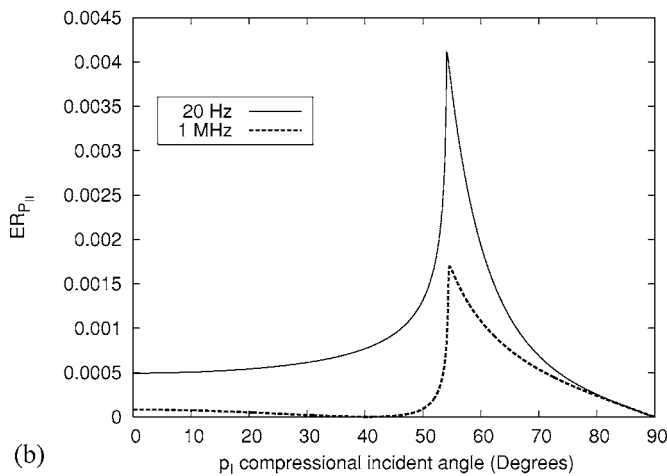
FIG. 1. Magnitude of the energy reflection coefficients at low and high frequencies for (a) p_{I} and (b) p_{II} compressional waves.

FIG. 2. Magnitude of the energy reflection coefficients at low and high frequencies for (a) s_{I} and (b) s_{II} shear waves.

Rubino *et al.*: Reflection and transmission in composite media
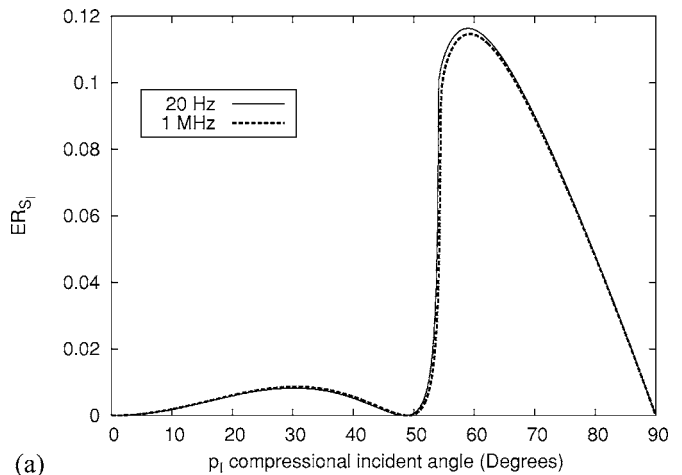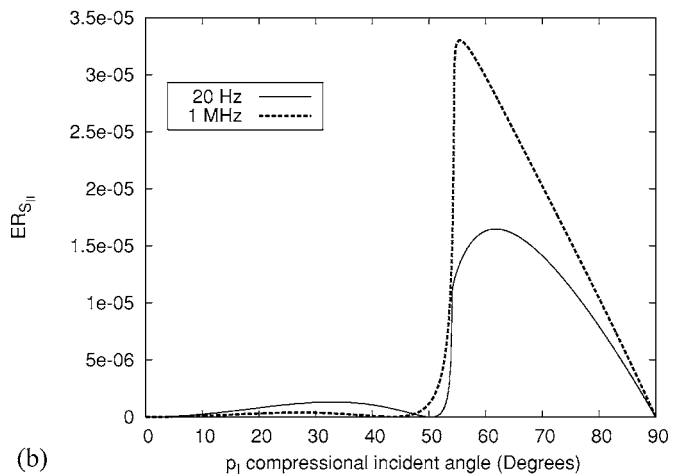
FIG. 3. Magnitude of the energy transmission coefficients at low and high frequencies for (a) p_I and (b) p_{II} compressional waves.
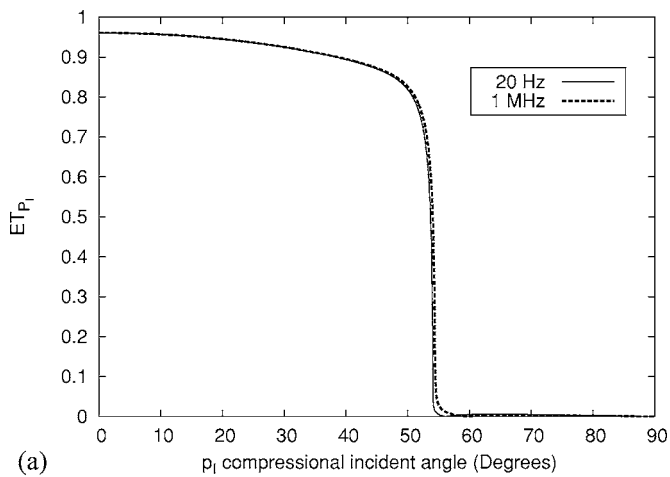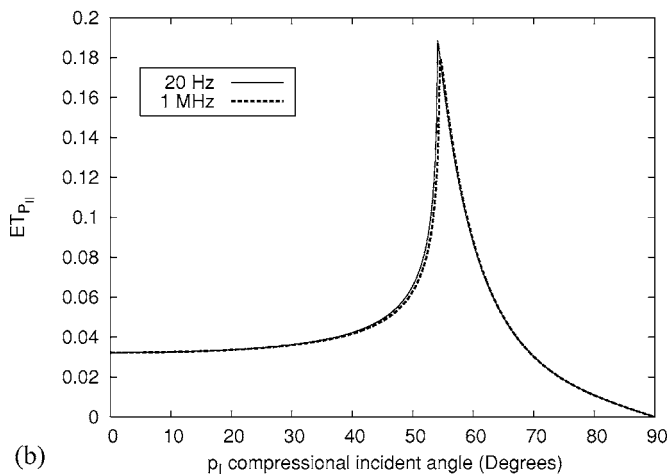
FIG. 4. Magnitude of the energy transmission coefficients at low and high frequencies for (a) s_I and (b) s_{II} shear waves.

observed in Figs. 1(b) and 2(b), the energy of the reflected slow waves is not significant for the present experiment.

In Figs. 2(a) and 4(a) it can be observed that for angles of incidence around the critical angle there is a significant conversion of incident energy into reflected and transmitted fast shear waves. For these angles of incidence, the energy converted into transmitted s_{II} shear wave is more than 1%, as observed in Fig. 4(b).

Tables III and IV present numerical results for the energy fluxes computed as described in Eqs. (4.11) and (4.12) for 20 Hz and 1 MHz, for a fixed angle of incidence equal to 45°. Note that in this case the interference fluxes are negligible with respect to the incident flux. Also, according to Eq. (4.16), the value $(F_t - F_r)$ at the bottom of Table IV is equal to F_{ip_I} (at the bottom of Table III), as expected.

VI. CONCLUSIONS

In this work the reflection and transmission coefficients for plane elastic waves arriving at a plane interface separating two porous composite half-spaces were defined and analyzed. The medium was assumed to be composed by two weakly coupled solids saturated by a single-phase fluid, according to the model proposed by Santos *et al.*[8] The energy flux Umov-Poynting vector appropriate for this kind of me-

dia was defined aiming to obtain the energy reflection and transmission coefficients and to check the corresponding amplitude coefficients as well.
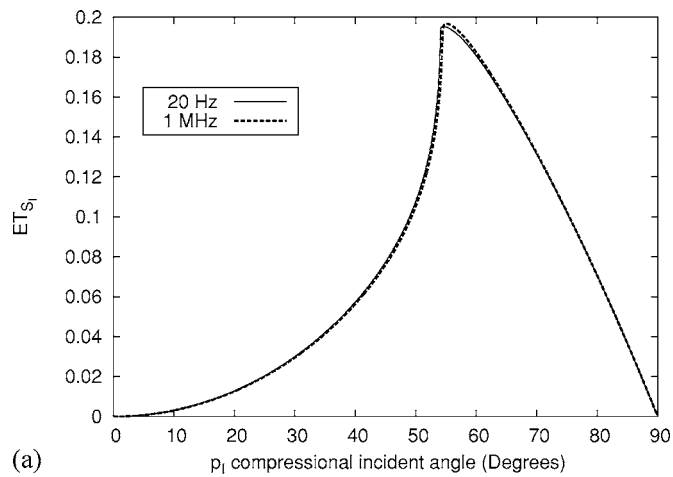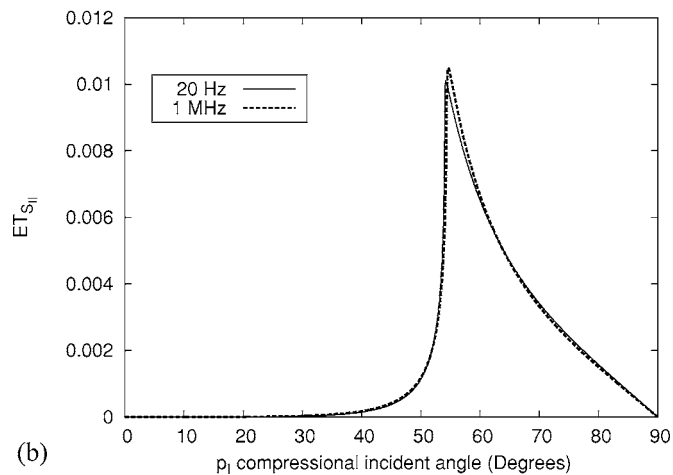
These calculations were applied to study the energy splitting at a plane interface defined by a change in ice content in the pores within a sample of water saturated partially frozen Berea sandstone. The same procedure can be applied to study similar problems in shaley-sandstones, gas-hydrate bearing sediments, or other composite porous media, which will be the subject of forthcoming publications.

The numerical example showed that, in some cases, an important part of the incident energy may be converted into slow waves energy. This fact allowed us to conclude that

TABLE III. Incidence energy fluxes (W/m^2) for two frequencies. The angle of incidence is 45°.

	Freq=20 Hz	Freq=1 GHz
F_{ip_I, ip_I}	$-51.224\,649\,3$	$-3.024\,657\,47E+20$
F_{ip_I, rp_I}	$-2.123\,951\,78E-05$	$-2.730\,194\,29E+15$
$F_{ip_I, rp_{II}}$	$3.793\,340\,36E-05$	$1.907\,449\,45E+15$
$F_{ip_I, rp_{III}}$	$-7.929\,781\,19E-06$	$8.412\,621\,63E+14$
F_{ip_I, rs_I}	$1.382\,501\,8E-05$	$-2.681\,281E+14$
$F_{ip_I, rs_{II}}$	$1.039\,111\,87E-08$	$1.610\,622\,46E+13$
F_{ip_I}	$-51.224\,626\,7$	$-3.024\,659\,81E+20$

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Rubino *et al.*: Reflection and transmission in composite media   2433

TABLE IV. Reflection and transmission energy fluxes (W/m^2) for two frequencies. The angle of incidence is 45°.

	Freq=20 Hz	Freq=1 MHz		Freq=20 Hz	Freq=1 MHz
F_{rp_I,rp_I}	0.002 775 754 06	5.495 216 42E+16	F_{tp_I,tp_I}	−44.586 300 9	−2.639 224 45E+20
$F_{rp_{II},rp_{II}}$	0.047 911 623 3	3.893 246 28E+15	$F_{tp_I,tp_{II}}$	−2.519 184 2	−1.454 391 6E+19
$F_{rp_{III},rp_{III}}$	0.000 407 264 28	1.297 357 35E+17	$F_{tp_{III},tp_{III}}$	−2.497 444 32E−06	−7.869 188 87E+16
F_{rs_I,rs_I}	0.085 152 149 6	6.231 091 46E+17	F_{ts_I,ts_I}	−3.963 762 83	−2.298 747 05E+19
$F_{rs_{II},rs_{II}}$	2.884 038 89E−05	1.584 913 93E+13	$F_{ts_{II},ts_{II}}$	−0.019 330 421 7	−1.280 725 14E+17
$F_{rp_I,rp_{II}}$	−7.584 280 38E−07	7.942 922 02E+13	$F_{tp_I,tp_{II}}$	0.000 226 297 985	−2.063 252 6E+16
$F_{rp_I,rp_{III}}$	−4.908 533 51E−08	4.943 405 55E+13	$F_{tp_I,tp_{III}}$	−8.941 580 64E−07	1.532 097 6E+15
F_{rp_I,rs_I}	−3.075 399 88E−07	−8.583 528 27E+13	F_{tp_I,ts_I}	−7.314 388 98E−05	−4.565 703 22E+15
$F_{rp_I,rs_{II}}$	5.410 540 61E−11	−2.540 749 69E+11	$F_{tp_I,ts_{II}}$	−1.853 187 8E−07	3.217 747 47E+15
$F_{rp_{II},rp_{III}}$	−2.139 559 67E−05	6.455 610 63E+14	$F_{tp_{II},tp_{III}}$	4.365 625 56E−07	1.066 102 84E+16
F_{rp_{II},rs_I}	8.093 845 98E−07	−2.815 880 21E+14	F_{tp_{II},ts_I}	5.754 615 64E−05	1.043 491 17E+16
$F_{rp_{II},rs_{II}}$	3.537 462 83E−10	−2.238 822 82E+11	$F_{tp_{II},ts_{II}}$	−1.392 591 69E−06	8.041 028 29E+15
F_{rp_{III},rs_I}	3.409 409 97E−07	−3.273 194 92E+13	F_{tp_{III},ts_I}	−1.195 371 23E−07	2.543 082 68E+15
$F_{rp_{III},rs_{II}}$	−5.623 850 68E−10	2.617 562 81E+11	$F_{tp_{III},ts_{II}}$	3.082 060 98E−09	6.811 625 39E+14
$F_{rs_I,rs_{II}}$	−1.716 981 86E−11	−1.585 814 92E+12	$F_{ts_I,ts_{II}}$	−1.404 094 69E−07	−5.218 874 37E+15
F_r	0.136 254 271	8.120 786 08E+17	F_t	−51.088 372 5	−3.0165 390 2E+20
			$F_t - F_r$	−51.224 626 7	−3.024 659 81E+20

disregarding the presence of slow waves may produce significant errors in the computation and analysis of reflection and transmission coefficients for frequencies lying between the seismic and ultrasonic ranges.

APPENDIX A: COMPUTATION OF THE COEFFICIENTS IN THE STRESS-STRAIN RELATIONS

1. The elastic coefficients

Let $K_{s1,m}, K_{s3,m}, \mu_{s1,m}$, and $\mu_{s3,m}$ denote the bulk and shear moduli of the two solid (dry) frames, respectively. Also, let $K_{s1}, \mu_{s1}, K_{s3}, \mu_{s3}$ denote the bulk and shear moduli of the grains in the two solid phases, respectively, and K_f, η the bulk modulus and viscosity of the fluid phase. Then, the elastic coefficients are given by[6,8]

$$\mu_j = [(1-g_j)\phi_j]^2 \mu_{av} + \mu_{sj,m}, \quad g_j = \frac{\mu_{sj,m}}{\phi_j \mu_{sj}}, \quad j=1,3,$$

$$\mu_{13} = (1-g_1)(1-g_3)\phi_1 \phi_3 \mu_{av},$$

$$\mu_{av} = \left[\frac{(1-g_1)\phi_1}{\mu_{s1}} + \frac{\phi}{2\omega\eta} + \frac{(1-g_3)\phi_3}{\mu_{s3}} \right]^{-1},$$

$$K_{av} = \left[(1-c_1)\frac{\phi_1}{K_{s1}} + \frac{\phi}{K_f} + (1-c_3)\frac{\phi_3}{K_{s3}} \right]^{-1},$$

$$c_j = \frac{K_{sj,m}}{\phi_j K_{sj}}, \quad j=1,3,$$

$$K_{Gj} = K_{sj,m} + (\alpha_j)^2 K_{av}, \quad \alpha_j = S_j - \frac{K_{sj,m}}{K_{sj}},$$

$$B_1 = \frac{S_1 \phi^2 K_{av} + C_{12}}{\phi}, \quad B_2 = \frac{S_3 \phi^2 K_{av} + C_{23}}{\phi},$$

$$B_3 = (C_{13} + S_3 C_{12} + S_1 C_{23} + S_1 S_3 \phi^2 K_{av}),$$

where

$$C_{12} = \phi K_{av}(\alpha_1 - S_1 \phi), \quad C_{23} = \phi K_{av}(\alpha_3 - S_3 \phi),$$

$$C_{13} = K_{av}(\alpha_1 - S_1 \phi)(\alpha_3 - S_3 \phi).$$

The moduli K_{G1} and K_{G3} are the analogues of Gassmann's moduli,[15] while the coefficients α_1 and α_3 correspond to the effective stress coefficients in the classic Biot theory.[10,16]

2. Viscoelastic moduli

Viscoelastic dissipation was introduced by using the correspondence principle stated by Biot,[16,17] in which the real poroelastic coefficients in the constitutive relations are replaced by complex frequency dependent poroviscoelastic moduli satisfying the same relations as in the elastic case. Among the different existing models, the linear viscoelastic model presented in Liu et al.[18] was chosen to make the moduli $K_{G1}, K_{G3}, \mu_1, \mu_3$ in Eq. (2.3) complex and frequency dependent. The set of poroviscoelastic moduli is computed using

$$M(\omega) = \frac{M_{re}}{R_M(\omega) - iT_M(\omega)},$$

where $M = K_{G1}, K_{G3}, \mu_1, \mu_3$ and the coefficient M_{re} is the relaxed elastic modulus associated with M.[19] The frequency dependent functions R_M and T_M, associated with a continuous spectrum of relaxation times, characterize the viscoelastic behavior and are given by[18]

$$R_M(\omega) = 1 - \frac{1}{\pi \hat{Q}_M} \ln \frac{1 + \omega^2 T_{1,M}^2}{1 + \omega^2 T_{2,M}^2},$$

$$T_M(\omega) = \frac{2}{\pi \hat{Q}_M} \tan^{-1} \frac{\omega(T_{1,M} - T_{2,M})}{1 + \omega^2 T_{1,M} T_{2,M}}.$$

The model parameters $\hat{Q}_M$, $T_{1,M}$, and $T_{2,M}$ are taken such that the quality factors $Q_M(\omega) = T_M/R_M$ are approximately equal to the constant $\hat{Q}_M$ within a wide frequency band, ranging from the low frequencies used in seismic exploration to the very high values used in ultrasonic testing. Values of $\hat{Q}_M$ range from $\hat{Q}_M = 20$ for highly dissipative materials to about $\hat{Q}_M = 1000$ for almost elastic ones.

APPENDIX B: MASS AND FRICTION COEFFICIENTS

Let ρ_m, $m = 1, 2, 3$, denote the mass density of each solid and fluid constituent in Q. Santos et al.[8] have taken the mass coupling coefficients to be real and frequency independent with $p_{jk} = m_{jk}$ and the m_{jk}-entries defined by

$$m_{11} = \rho_2 \phi (1 + (S_1)^2 a_{32} + (S_3)^2 a_{12} - 2S_3 - (S_1)^2)$$

$$+ a_{13} \rho_1 \phi_1 + (a_{31} - 1) \rho_3 \phi_3,$$

$$m_{12} = \rho_2 (S_3(1 - a_{12}) + S_1 a_{32}),$$

$$m_{13} = \rho_2 \phi (1 - (S_1)^2 a_{32} - (S_3)^2 a_{12} - S_1 S_3)$$

$$+ \rho_1 \phi_1 (1 - a_{13}) + \rho_3 \phi_3 (1 - a_{31}),$$

$$m_{22} = \frac{\rho_2}{\phi}(a_{12} + a_{32} - 1),$$

$$m_{23} = \rho_2 (S_1(1 - a_{32}) + S_3 a_{12}),$$

$$m_{33} = \rho_2 \phi (1 + (S_1)^2 a_{32} + (S_3)^2 a_{12} - 2S_1 - (S_3)^2)$$

$$+ a_{31} \rho_3 \phi_3 + (a_{13} - 1) \rho_1 \phi_1.$$

Here,

$$a_{12} = \frac{\phi_1 \rho}{\phi \rho_2} r_{12} + 1, \quad a_{32} = \frac{\phi_3 \rho'}{\phi \rho_2} r_{32} + 1,$$

with the r_{jk}'s being the geometrical aspects of the boundaries separating the phases j and k (equal to $1/2$ for spheres). It is assumed that[7] $a_{13} = a_{31} = 1$, $r_{12} = r_{32} = 1/2$, and

$$\rho = \frac{\phi \rho_2 + \phi_3 \rho_3}{\phi + \phi_3}, \quad \rho' = \frac{\phi \rho_2 + \phi_1 \rho_1}{\phi + \phi_1}.$$

Next, the coefficients f_{ij}'s were computed as follows.[8] For the case of frozen porous media, let

$$d_{12} = \frac{\phi^2}{\kappa_1}, \quad d_{23} = \frac{\phi^2}{\kappa_3},$$

where the permeability coefficients κ_1 and κ_3 are defined in terms of the absolute permeabilities $\kappa_{1,0}$ and $\kappa_{3,0}$ of the two solid frames by (see also the works of Leclaire et al.[5,9])

$$\kappa_1 = \kappa_{1,0} \frac{\phi^3}{(1 - \phi_1)^3}, \quad \kappa_3 = \kappa_{3,0} \frac{(1 - \phi_1)^2}{(\phi_3)^2} \left(\frac{\phi}{\phi_1}\right)^3.$$

Then the frequency independent friction coefficients f_{11}, f_{12}, and f_{22} are determined from

$$g_{11} = d_{12}(S_3)^2 + d_{23}(S_1)^2, \quad g_{12} = \frac{d_{12}S_3 - d_{23}S_1}{\phi},$$

$$g_{22} = \frac{d_{12} + d_{23}}{\phi^2},$$

$$f_{11} = \eta g_{11} + g_{13}, \quad f_{12} = \eta g_{12}, \quad f_{22} = \eta g_{22}. \quad (B1)$$

Since the coefficient g_{13} takes into account friction between the two solid frames, a proper model based, for example, in a Coulomb-type friction theory may be used, but for simplicity in all the numerical examples presented in this article the coefficient g_{13} was set to be zero.[6,8] This is consistent with the thermodynamical condition stated by Santos et al.[8] [Eq. (B9)] about the non-negative character of the dissipation function.

In the high frequency range the set of inertial and friction coefficients m_{jk}, $1 \le j, k \le 3$, and f_{11}, f_{12}, and f_{22} need to be modified to include the departure of the relative flow from laminar type above a certain critical frequency depending on the pore radius as explained by Biot.[17] For that purpose, the dissipative terms f_{11}, f_{12}, and f_{22} are modified by multiplying the fluid viscosity η by the frequency correction factor $F_R(\theta)$, where the complex valued frequency dependent function $F(\theta) = F_R(\theta) + iF_I(\theta)$ is the frequency correction function defined by Biot:[17]

$$F(\theta) = \frac{1}{4} \frac{\theta T(\theta)}{1 - \frac{2}{i\theta} T(\theta)}, \quad T(\theta) = \frac{\text{ber}'(\theta) + i \, \text{bei}'(\theta)}{\text{ber}(\theta) + i \, \text{bei}(\theta)},$$

with $\text{ber}(\theta)$ and $\text{bei}(\theta)$ being the Kelvin functions of the first kind and zero order. The frequency dependent argument $\theta = \theta(\omega)$ is given in terms of the pore size parameter a_p by

$$\theta = a_p \sqrt{\omega \rho_2 / \eta}, \quad a_p = 2\sqrt{\kappa A_0 / \phi},$$

where $1/\kappa = 1/\kappa_1 + 1/\kappa_3$ and A_0 is the Kozeny-Carman constant.[20,13]

Consequently the frequency dependent mass and viscous coupling coefficients are defined in the following fashion:

$$p_{11}(\omega) = m_{11} + \frac{\eta F_I(\theta) g_{11}}{\omega}, \quad p_{12}(\omega) = m_{12} - \frac{\eta F_I(\theta) g_{12}}{\omega},$$
$$(B2)$$

$$p_{13}(\omega) = m_{13} - \frac{\eta F_I(\theta) g_{11}}{\omega}, \quad p_{22}(\omega) = m_{22} + \frac{\eta F_I(\theta) g_{22}}{\omega},$$
$$(B3)$$

$$p_{23}(\omega) = m_{23} + \frac{\eta F_I(\theta) g_{12}}{\omega}, \quad p_{33}(\omega) = m_{33} + \frac{\eta F_I(\theta) g_{11}}{\omega},$$
$$(B4)$$

$$f_{11}(\omega) = \eta F_R(\theta)g_{11} + g_{13}, \tag{B5}$$

$$f_{12}(\omega) = \eta F_R(\theta)g_{12}, \quad f_{22}(\omega) = \eta F_R(\theta)g_{22}.$$

[1] H. Deresiewicz and J. T. Rice, "The effect of boundaries on wave propagation in liquid filled porous solids I. Reflection of plane waves at a true plane boundary," Bulletin of the Seismological Society of America, **52**, 627–638 (1960).

[2] N. C. Dutta and H. Odé, "Seismic reflections from a gas-water contact," Geophysics **48**, 148–162 (1983).

[3] J. E. Santos, J. M. Corbero, C. L. Ravazzoli, and J. L. Hensley, "Reflection and transmission coefficients in fluid-saturated porous media," J. Acoust. Soc. Am. **91**, 1911–1923 (1992).

[4] A. I. M. Denneman, G. G. Drijkoningen, D. M. J. Smeulders, and K. Wapenaar, "Reflection and transmission of waves at a fluid/porous-medium interface," Geophysics **67**, 282–291 (2002).

[5] Ph. Leclaire, F. Cohen-Tenoudji, and J. Aguirre Puente, "Extension of Biot's theory of wave propagation to frozen porous media," J. Acoust. Soc. Am. **96**, 3753–3767 (1994).

[6] J. M. Carcione and U. Tinivella, "Bottom-simulating reflectors: Seismic velocities and AVO effects," Geophysics **65**, 54–67 (2000).

[7] J. M. Carcione, J. E. Santos, C. L. Ravazzoli, and H. B. Helle, "Wave simulation in partially frozen porous media with fractal freezing conditions," J. Appl. Phys. **94**, 7839–7847 (2003).

[8] J. E. Santos, C. L. Ravazzoli, and J. M. Carcione, "A model for wave propagation in a composite solid matrix saturated by a single-phase fluid," J. Acoust. Soc. Am. **115**, 2749–2760 (2004).

[9] Ph. Leclaire, F. Cohen-Tenoudji, and J. Aguirre Puente, "Observation of two longitudinal and two transverse waves in a frozen porous medium," J. Acoust. Soc. Am. **97**, 2052–2055 (1995).

[10] J. M. Carcione, *Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic and Porous Media*, Handbook of Geophysical Exploration, Vol. 31 (Pergamon, New York, 2001).

[11] V. Cerveny, "Energy flux of time-harmonic waves in anisotropic dissipative media," in Seismic waves in complex 3-D structures, Consorptium Research Project, Dept. of Geophysics, Charles University, Prague, Report No. 11, 2001.

[12] G. T. Kuster and M. N. Toksöz, "Velocity and attenuation of seismic waves in two-phase media. 1. Theoretical formulations," Geophysics **39**, 587–606 (1974).

[13] J. M. Hovem and G. D. Ingram, "Viscous attenuation of sound in saturated sand," J. Acoust. Soc. Am. **66**, 1807–1812 (1979).

[14] E. G. Henneke II, "Reflection-refraction of a stress wave at a plane boundary between anisotropic media," J. Acoust. Soc. Am. **51**, 210–217 (1971).

[15] F. Gassmann, "Über die elastizität pöroser medien" ("On the elasticity of porous media"), Vierteljahrsschr. Natforsch. Ges. Zur. **96**, 1–23 (1951).

[16] M. A. Biot, "Mechanics of deformation and acoustic propagation in porous media," J. Appl. Phys. **33**, 1482–1498 (1962).

[17] M. A. Biot, "Theory of deformation of a porous viscoelastic anisotropic solid," J. Appl. Phys. **27**, 459–467 (1956).

[18] H. P. Liu, D. L. Anderson, and H. Kanamori, "Velocity dispersion due to anelasticity; implications for seismology and mantle composition," Geophys. J. R. Astron. Soc. **147**, 41–58 (1976).

[19] T. Bourbie, O. Coussy, and B. Zinszner, *Acoustic of Porous Media* (Editions Technip, Paris, 1987).

[20] J. Bear, *Dynamics of Fluid in Porous Media* (Dover, New York, 1972).

A comparison of sediment reflection coefficient measurements to elastic and poro-elastic models

H. John Camin[a)]
Applied Research Laboratory, The Pennsylvania State University, State College, PA 16804–0030

Marcia J. Isakson[b)]
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas

This work compared three plane wave reflection coefficient models to laboratory sediment reflection measurements collected using a spherical source and receiver. Plane wave decomposition was used to modify the reflection coefficient models to include the inherent spherical effects such as lateral wave interference for direct comparison with the data. Sets of data at two transducer separation distances (1.27 and 0.25 m) were collected to compare the range dependence of the spherical effects. Bandpass filtered linear frequency-modulated chirps from 30 to 160 kHz were used to measure frequency dependence. Grazing angles from 5° to 75° were measured to compare angle dependence. Each set of data was collected along approximately 3 m of smoothed sediment for spatial averaging. Unavoidable experimental effects including transducer response, beam pattern, and spherical spreading were accounted for in order to compare the reflection coefficient measurements with the modified models. Significant spherical wave effects were measured in the data. Three reflection coefficient models were considered: the viscoelastic model, the grain shearing model [M. Buckingham, J. Acoust. Soc. Am. **108**, 1796–2815 (2000)] and the effective density fluid model [K. Williams, J. Acoust. Soc. Am. **110**, 2276–2281 (2001)]. The viscoelastic and grain shearing models predicted values for the reflection coefficient that were not within the 95% confidence interval for low frequencies. The data exhibited high variance which was frequency and angle dependent. This variance is not likely to be caused by variations in bulk properties as defined by the fluid or viscoelastic models. The cause of the variance will be considered in subsequent publications. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354002]

I. INTRODUCTION

There is an increasing need in ocean acoustics to fully understand acoustic wave interaction with sediments. As more and more applications are focusing on shallow water environments, the importance of accurate sediment models is ever increasing. The plane wave reflection coefficient is particularly important in underwater acoustics. Once the reflection coefficient is known, realistic propagation modeling can assist in communication, detection, and classification in littoral environments. Several plane wave models exist that are used to calculate the reflection coefficient given the sediment material parameters. These plane wave models are effective in the majority of situations encountered in real-world activities.

Spherical wave sources present a unique challenge as the plane wave models cannot properly account for the effects of a very broad beam pattern. Using spherical sources in long range applications is often a nonissue since a spherical wave eventually becomes locally planar. Close range applications, on the other hand, are affected greatly by the spherical wave effects. This is particularly important in a laboratory environment, where longer ranges are not feasible.

Measuring acoustic observables in a laboratory environment has many benefits, though. Laboratory measurements can be used to discern which models are most accurate; and assuming an accurate model, they can be inverted to determine well characterized sediment properties. Ideally, the laboratory is a carefully controlled environment where experimental parameters like incident angle, surface roughness and location are all precisely regulated. *In situ* measurements have increased variability that makes comparisons to theoretical models often difficult. In order to compare laboratory measurements made with spherical sources to the plane wave models, the models must first be modified to include the spherical wave effects.

It is a misnomer to call measurements of the bistatic response at the specular angle reflection coefficient measurements when the data are modified by spherical wave effects. However, these measurements will be called, "reflection coefficient measurements" throughout the text for clarity. When spherical wave effects are included in the reflection coefficient, the nomenclature, "spherical wave reflection coefficient" will be explicitly used.

---
a)Electronic address: hjc130@psu.edu
b)Electronic address: misakson@arlut.utexas.edu

The work presented here compares experimental spherical wave reflection coefficient measurements to three common plane wave reflection coefficient models that have been modified to include the spherical wave effects. A wide range of frequencies and grazing angles (30–160 kHz, 5°–75°) were tested at two ranges (1.27 and 0.25 m) to accurately compare experimental measurements to predicted model values. The three models that are compared are the viscoelastic model,[1] Buckingham's grain shearing model,[2] and Williams effective density fluid model (EDFM).[3] The elastic model assumes the sediment has a homogeneous composition and the averaged or bulk parameters for the wet sediment are used. The grain shearing and effective density models are poroelastic models that assume the individual sediment grains interact separately from the surrounding fluid.

The three models were used in plane wave decomposition (PWD) analysis[1] to include the spherical wave effects. PWD represents the spherical wave as a summation of plane waves which are individually modified by the respective plane wave reflection coefficients. The resulting summation which inherently takes into account all spherical wave effects can be directly compared to the measured data. The models and the PWD method will be further discussed in Sec. II.

The experiments were conducted during the summer of 2004 at the Applied Research Laboratories, The University of Texas at Austin. The experimental setup is discussed in Sec. III. Data analysis procedures are presented in Sec. IV. Experimental effects are discussed in Sec. V and results including the comparison of the data to models are shown in Sec. V. Finally, in Sec. VI, conclusions are drawn from the data/model comparisons.

II. THEORY

A. Viscoelastic model

The viscoelastic reflection coefficient model is a function of the incident angle and the characteristic acoustic impedances of the two mediums. The model includes an additional impedance for the shear wave motion. For the viscoelastic case[4]

$$R = \frac{Z_{tot} - Z_1}{Z_{tot} + Z_1},$$
(1)

where

$$Z_1 = \frac{\rho_1 c_1}{\cos \theta_i},$$
(2)

$$Z_{tot} = Z_c \cos^2 2\gamma_s + Z_s \sin^2 2\gamma_s,$$
(3)

$$Z_c = \frac{\rho_2 \omega}{(k_c^2 - k_1^2 \sin^2 \theta_i)^{\frac{1}{2}}},$$
(4)

$$Z_s = \frac{\rho_2 \omega}{(k_s^2 - k_1^2 \sin^2 \theta_i)^{\frac{1}{2}}}.$$
(5)

In the above equations, R is the reflection coefficient and Z_1 is the impedance of the water column. Z_{tot} is the total impedance of the sediment and Z_c is the impedance of the com-

pressional wave while k_c is the complex compressional wave number. Z_s is the impedance of the shear wave while k_s is the complex shear wave number and ω is the radial frequency. The viscoelastic model is completely defined by seven parameters and the incident angle, θ_i: the sound speed and density of the water column, c_1 and ρ_1, the sound speed of the compressional and shear waves, c_p and c_s, the density of the second medium, ρ_2, and the compressional and shear attenuations. The transmission angle of the shear wave, γ_s can be determined using Snell's law.

Equation (1) predicts a critical angle based on the ratio of the sound speeds both compressional and shear. However, if the shear wave speed in the sediment is much lower than the sound speed in the upper fluid, there will not be a second critical angle. Additionally, the Z_s term would then be negligible compared to the other impedances. In this case, the elastic reflection coefficient reduces to the fluid reflection coefficient.

B. Effective density fluid model

The sediment can also be considered as a poroelastic medium in which the displacement of the fluid and the displacement of the sand grains are considered independently. In a poroelastic model, there are three possible waves: a compressional wave in which the fluid and sand grain frame are moving in phase with each other, a compressional wave in which the fluid and sand grain frame are out of phase and a shear wave. The out-of-phase compressional wave is referred to colloquially as the "slow" wave. One formulation of a poro-elastic model is the Biot/Stoll model.[5–7] The Biot/Stoll model has 13 independent parameters, several of which are difficult to measure directly. The parameters are classed into three groups: bulk parameters, fluid motion parameters, and frame parameters. The bulk parameters are porosity β, pore size parameter α, grain density ρ_g, fluid density ρ_f, grain bulk modulus K_g, and fluid bulk modulus K_f. The fluid motion parameters include viscosity η, permeability κ and tortuosity τ. Last, in Biot/Stoll theory the sand grains are assumed to form a frame with its own complex bulk and shear moduli. The frame moduli are very difficult to measure, and they are generally assumed to be much smaller than the bulk moduli. If the frame parameters are assumed to zero, the Biot/Stoll equations can be reformulated into fluid-like equations.[3] The poroelastic effects are contained in the "effective density" term and a complex sound speed. The effective density is given by

$$\rho_{eff} = \rho_f \frac{\alpha(1-\beta)\rho_g + \beta(\alpha-1)\rho_f + \frac{i\beta\rho F\eta}{\rho_f \omega \kappa}}{\beta(1-\beta)\rho_g + (\alpha - 2\beta + \beta^2)\rho_f + \frac{i\beta F\eta}{\omega \kappa}},$$
(6)

where ρ is the sediment density given by $\beta\rho_f + (1-\beta)\rho_g$ and F is a function derived by Biot to account for the deviation from Poisuille flow at higher frequencies.[7] The complex sound speed is represented by

Camin and Isakson: Comparison of spherical wave reflection measurements

$$c = \sqrt{\frac{\left(\frac{1-\beta}{K_g} + \frac{\beta}{K_f} \right)^{-1}}{\rho_{\text{eff}}}}. \tag{7}$$

The resulting complex density from Eq. (6), compressional sound speed from Eq. (7) and attenuation which given by the imaginary part of ω/c from Eq. (7) are used in the fluid model described above.

C. Grain shearing model

Recently, Buckingham has suggested that acoustic reflection phenomena can also be modeled by considering the physics of the grain contact.[2] Translational and rotational shearing of the individual grains against one another amounts to a "random stick-slip" motion statistically combined to produce a relaxation mechanism. Using the average density, the grain shearing model yields the shear and compressional sound speeds and the shear and compressional attenuations, which are input into the elastic model detailed above. According to Buckingham, the equation for the bulk density of the saturated sediment is based on the porosity β and fluid and grain densities ρ_f, ρ_s:

$$\rho_2 = \beta \rho_f + (1-\beta)\rho_s. \tag{8}$$

The equations for the compressional and shear sound speeds c_p, c_s and attenuations α_p, α_s are given by

$$\frac{1}{c_p} = \frac{1}{c_0}\text{Re}\left[\left(1 + \frac{3\gamma_p + 4\gamma_s}{3\rho c_0^2}(i\omega)^n \right)^{-1/2} \right], \tag{9}$$

$$\alpha_p = -\frac{\omega}{c_0}\text{Im}\left[\left(1 + \frac{3\gamma_p + 4\gamma_s}{3\rho c_0^2}(i\omega)^n \right)^{-1/2} \right], \tag{10}$$

$$\frac{1}{c_s} = \frac{1}{c_0}\sqrt{\frac{\rho c_0^2}{\gamma_s}}|\omega|^{-n/2}\cos\left(\frac{n\pi}{4} \right), \quad \text{and} \tag{11}$$

$$\alpha_s = \frac{|\omega|}{c_0}\sqrt{\frac{\rho c_0^2}{\gamma_s}}|\omega|^{-n/2}\sin\left(\frac{n\pi}{4} \right). \tag{12}$$

Here ρ is the sediment density defined in Eq. (8). The compressional wave speed in an equivalent suspension, c_0, is determined by Wood's equations.[8] The radial stress-relaxation exponent, n, is computed from initial empirical measurements. The compressional and shear rigidity coefficients, γ_p and γ_s, are solved simultaneously from empirical measurements as well.

D. Plane wave decomposition

Since sediments have an angular dependent reflection coefficient which varies in amplitude and phase, the reflection from a flat sediment surface will not produce a perfectly spherical reflection, but a distorted semispherical wave. Furthermore, if the sediment has a higher sound speed than the fluid above, there is a critical angle past which there is total internal reflection. At this critical angle, the acoustic wave propagates along the fluid/sediment interface at the speed of sound in the sediment and reradiates back into the fluid at the same critical angle. This phenomenon is called the head
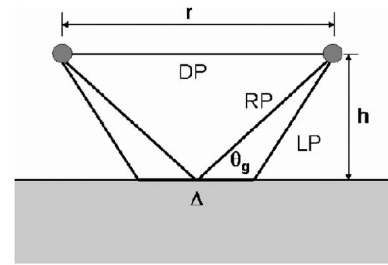
FIG. 1. Ray paths of the direct wave, specular reflection, and lateral wave.

wave or lateral wave since it travels ahead of the reflected acoustic wave in the water. Figure 1 shows the various ray paths between a source and a receiver that are possible with spherical waves. The interested reader can find color pictorials of lateral waves in Ref. 9.

From the figure, the direct path (DP) is the shortest distance between a source and a receiver. The specular reflection path (RP) is incident on the sediment at a grazing angle dependent on the position of the source and receiver. The lateral wave path represents an additional ray that may cause interference with specular reflection. It is simplistic to consider only the ray theory approximation of the lateral wave when addressing spherical effects since, around the critical angle, ray theory will produce infinite intensity at caustics, points in space where rays cross. In order to calculate the spherical wave reflection coefficient, the evaluation of the entire field integral is necessary.[10] Also, it is important to note that spherical wave effects will only be prevalent for source receiver ranges of less than 100–200 wavelengths and only for small grazing angles. For example, at 140 kHz in seawater, spherical effects are negligible after about 2 m, but at 40 kHz, spherical effects are prevalent up to about 5–7 m subcritically. This is due to attenuation and spreading of the lateral wave.

Section 26 in Brekhovskikh[1] details the development of a method to represent a spherical wave as an infinite summation of plane waves. The plane wave reflection coefficients computed from various models can be linearly multiplied by the individual plane waves inside the integral. The resulting integration yields the spherical wave reflection coefficient. The first step in PWD is to assume the incoming wave has the form of a spherical wave in three dimensions where the time dependence ($e^{-i\omega t}$) has been suppressed and only relative pressures are considered

$$P_i = \frac{p_0 e^{ikr}}{r}, \tag{13}$$

where

$$r = \sqrt{x^2 + y^2 + z^2}. \tag{14}$$

The received pressure, P_r, is a function of the initial pressure, p_0, the wave number, k, and the linear distance between the source and receiver, r. The distance between the two is a function of the coordinate displacements, x, y, and z from the source. As shown in Eq. (15), the expansion into the Fourier space, or the Helmholtz equation, reduces to

$$P_r = p_0 \frac{i}{2\pi} \int \int_{-\infty}^{\infty} e^{i(k_x x + k_y y \pm k_z z)} \frac{dk_x dk_y}{k_z}. \qquad (15)$$

Brekhovskikh shows that by converting to polar coordinates and making use of several identities and substitutions the spherical wave reflection coefficient (R_{SW}) can be written in the form of

$$R_{SW} = \frac{p_0 i k}{P_i} \int_0^{\pi/2 - i\infty} R(\theta) J_0(kr \sin\theta) e^{i2kh\cos\theta} \sin\theta d\theta. \qquad (16)$$

Here h is the height above bottom for the transducer array, θ is the grazing angle, and P_i is the incident spherical wave. Equation (16) is the PWD integral that is used to modify the plane wave reflection models for comparison with spherical wave data. It yields the reflected pressure for a given geometry based on the height above the sediment and the range between the transducers. In order to modify a given plane wave reflection coefficient model, the plane wave reflection coefficient, $R(f, \theta_g)$, computed from the model of interest is linearly multiplied by the kernel of the PWD integration.

Although traditional analysis involves many simplifications of Eq. (16) and typically various forms of saddle point integration, current computer processing power facilitates the direct numerical integration in this work. This greatly simplifies the conceptual procedure.

In the past, numerical integration has successfully been used to verify Brekhovskikh's work.[11] Until recently, however, the computational requirements were too high to make this approach a feasible means of solving repeated integrals.

Numerically integrating Eq. (16) near normal incidence and near zero degrees grazing angle is difficult. The frequency of oscillation of the function is much higher at normal incidence than at other angles. In this region, a sufficiently fine grid size results in large computation time. As the glazing angle decreases, the exponential decay in the imaginary portion of the PWD kernel decreases and a sinusoidal component becomes dominant. The limit occurs at exactly zero degrees where the imaginary portion is a varying sinusoid to minus infinity. However, with the transducer geometry used in this experiment, it was impractical to measure data at grazing angles of more than about 70° grazing or less than about 7° grazing. Therefore, it was unnecessary to numerically calculate the spherical wave reflection coefficient at the limits.

Figure 2 shows the comparison of the numerical integration of PWD integral with the plane wave reflection coefficient for a fluid model. The results were computed at two different separation distances and frequencies to show the general trends. As suspected, lower frequencies and closer distances cause greater spherical effects. Note that even at 1.27 m apart and at 112 kHz, there is some indication of spherical wave effects.

III. EXPERIMENTAL SETUP

Reflection coefficient measurements were taken in the ARL:UT Sonar Modal Calibration Tank 1. The tank is
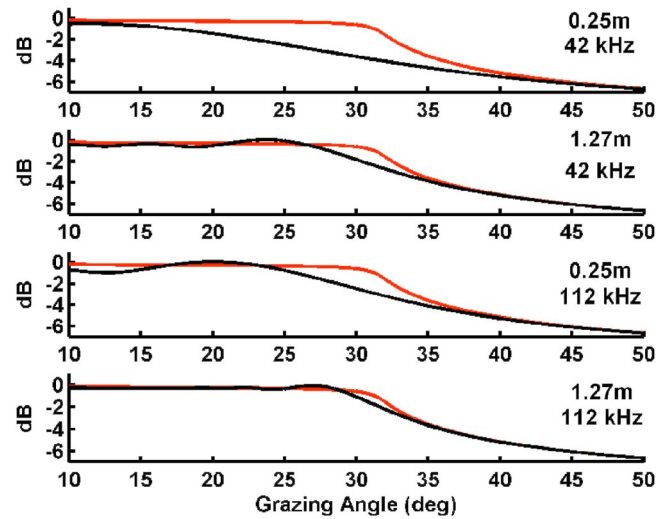
FIG. 2. PWD results (dark) compared with the plane wave reflection coefficient for the fluid model (light). The spherical waves cause significant changes in the apparent critical angle particularly at low frequencies and separation distances.

18.3 m long, 4.5 m wide, and 3.6 m deep. At the bottom there is 1 m of unwashed river sand that was collected from the lower Colorado River in the 1970s. It has remained generally undisturbed for over 30 years. Biological activity is kept to a minimum through periodic chlorination. The tank has a slight temperature gradient of 0.45 °C/m extending approximately 0.3 m below the air/water interface. Below 0.3 m depth, the temperature is constant. Situated above the tank are two moving platforms, each equipped with a vertical shaft capable of positioning 150 lb of equipment at any depth in the water column. It provides an ideal location to precisely measure the reflection coefficient of sediments as a function of grazing angle.

The sediment in the tank includes various in homogeneities commonly found in river sand. Though it is generally fine sand, there are numerous rocks, and chunks of clay. A 2002 study characterizing the sediment found the sound speed to be 1740 ± 20 m/s at 30 kHz.

The average total density was measured to be 1.99 g/cm³. A detailed grain analysis was performed in 1992. A visual analysis at that time revealed a majority of quartz crystals with some silt and feldspar. The majority of the grains were poorly sorted and angular to subangular in shape. The porosity was measured to be 0.40. The average grain size diameter is 0.25 mm based on sieve measurements.

Figure 3 shows an elevation schematic of Calibration Tank 1. The moving platform can horizontally traverse the entire length of the tank. Attached to the vertical column is a two transducer acoustic test frame. The frame itself is narrow in depth and can easily be reversed to take data both from the sediment/water interface or the air/water interface. The frame was engineered so that it had less than a 1 mm deflection in either configuration. For the sediment/water interface data collection at the larger range, a cork reflector was mounted at a 45 deg angle above the sending transducer to exclude the air/water surface reflection for the sediment data. (See Fig. 4.)
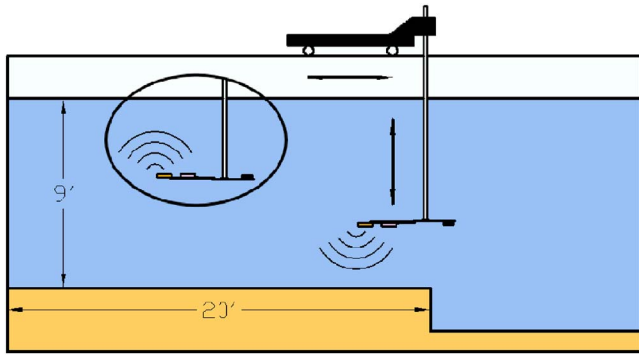
FIG. 3. ARL Sonar Model Tank with experiment frame (to scale). Inset shows air/water configuration.

FIG. 5. Schematic of experiment and signal processing layout.

Each transducer was a spherical ITC 1089 transducer positioned so that the *XY* planes of the transducers were aligned. The z axis points along the transducer pigtail. The beam patterns of the two spherical transducers are nominally constant to 1 dB about the *XY* plane.

The reflection coefficient as a function of angle was measured by changing the height of the transducer array to sample different angles. Figure 1 shows that a change in height above the interface directly corresponds to a change in the grazing angle, θ_g, assuming the range, r, is held constant. The individual rays indicate the arrivals via the direct path and reflected path. The direct path will experience normal spherical spreading and will decrease in amplitude at a rate of $1/r$. The reflected path will experience a similar spreading loss due to the added distance, $\cos \theta_g / r$ where θ_g is the grazing angle.

To insure spatial stability, approximately 30 different locations on the sand were sampled. At each location, 75 angle measurements were made by coherently averaging 300–500 individual pings together. A 1 ms linear frequency modulated chirp, from 30 to 160 kHz, was used to sample a range of frequencies with each ping. Bandpass filtering allowed 10 subbands to be analyzed in each ping. This method uses one measurement to take the place of ten measurements at indi-

FIG. 4. Underwater view of the experiment apparatus and smoothed sediment. A reflector was added above the sending transducer to deflect rays from interacting with the air/water interface.

vidual frequencies. The above procedure was conducted at two different ranges between the transducers to compare the spherical effects as a function of range. The collected data represent one of the most comprehensive acoustic analysis of the sediment in ARL:UTs calibration tank to date.

Rough interface scattering effects were minimized by smoothing the sand with a 4 in. PVC pipe filled with scrap steel. The weighted pipe was gently dragged across the bottom several times. The laser line scan method developed at ARL:UT could not detect any surface roughness. This method has millimeter accuracy for spatial wavelengths from 0.1 to 1 m.[12]

The average root-mean-square roughness of smoothed, dry sand taken from the calibration tank was measured to be approximately 0.5 mm. This is about twice the average grain size, but is not unexpected. Various effects can increase the apparent roughness such as imperfections in the smoothing process, adhesion between sand grains and grain size distribution. Roughly 62% of the grains by mass are larger than the average grain size diameter.[*] The packing order of the individual grain sizes may also affect the apparent roughness.

An experiment test frame was designed and built in order to provide a stable platform that accommodated both air/water interface and sediment/water interface measurements. The sediment reflection measurements were normalized by the air/water data in order to correct for beam pattern and transducer response effects. This method improves the accuracy of the transducer response to nominally 0.25 dB.

The entire experiment, broken down into transfer functions, is shown in Fig. 5. The analog to digital (ADC) and digital to analog (DAC) converters on a National Instruments™ PCI-6115 data acquisition card were configured to trigger off of the same 4 MHz clock. A linearly amplified

signal from the DAC is applied directly to the 1TC 1089 transmit transducer, T_x. This produces relatively low amplitude, frequency dependent sound pressure levels of 155–170 dB in the water column. Because the transducers deviated from an ideal spherical source, slight beam pattern effects, B, contributed to the overall signal. The received signal also includes effects from spherical spreading, S_p, and the spherical reflection from the interface, $R(f, \theta_g)$. The received signal (X_t), including both the direct path pressure (X_{DP}) and reflected pressure (X_{RP}), is modified by the receiving transducer response including amplification.

IV. SIGNAL PROCESSING

Measuring the spherical wave reflection coefficient required accounting for several additional effects, in addition to the spherical wave effects, that were inherent in the experiment.

Each of the physical steps in the experiment, as outlined in Sec. III, can be reduced to transfer functions for signal processing. These transfer functions operate on the Fourier transform of the pressure amplitude

$$\eta_{DP}(f, \theta_0) = \text{TR}(f)B(f, \theta_0)Sp(\theta_0), \tag{17}$$

$$\eta_{RP}(f, \theta_g) = \text{TR}(f)B(f, \theta_g)Sp(\theta_g)R(f, \theta_g). \tag{18}$$

In general, two signals can be isolated in the data, the direct path pressure amplitude and the received path pressure amplitude. The amplitude of the direct path (η_{DP}) includes the transducer response TR, the beam pattern B, and spreading S_p effects the direct path angle θ_0 which was zero degrees grazing. The transducer response includes all effects from the amplifiers and DA/AD conversions. The beam pattern effects include both source and receiver beam patterns. The amplitude of the reflected path (η_{RP}) includes the same effects as the direct path, but are dependent on the particular grazing angle of each measurement. In addition, the reflected path includes the reflection R effects associated with interaction with the sediment or the air/water interface. These simplifications allow the different received pressure amplitudes to be described as

$$X_i(f) = I(f) + N(f), \tag{19}$$

$$X_{DP}(f, \theta_0) = X_i(f)\eta_{DP}(f, \theta_0) + N_{DP}(f), \tag{20}$$

$$X_{RP}(f, \theta_g) = X_i(f)\eta_{RP}(f, \theta_g) + N_{RP}(f). \tag{21}$$

The input signal X_i includes only the initial wave form I with some small added electronic random noise $N(f)$. Both the direct path amplitude X_{DP} shown in Eq. (20) and the reflected path amplitude X_{RP} shown in Eq. (21) include the input signal, but have the additional effects associated with each. The respective noise terms in each equation represent the contribution of the noise in Eq. (19) as well as added random noise $N(f)$ associated with each path. For the direct path the noise is associated with local temperature variations in the water column and is expected to be small. Some possible noise contributions for the reflected path are discussed in Sec. V E. The sum of the direct and reflected paths was
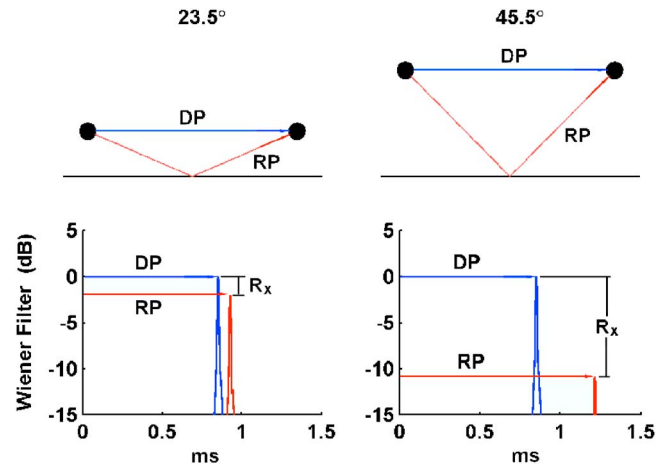
FIG. 6. Illustration of experiment and signal processing. The direct path time is set to zero before computing the grazing angle.

recorded. However, since the distance between the transducers is constant, the direct path contribution for any measurement is the same. Therefore, the direct and reflected paths can be separated by subtracting a temporally isolated direct path that was recorded in the middle of the water column before each set of measurements.

The process for calculating the reflection coefficient from a measured time series is as follows. First, the direct and reflected path arrival times were determined using a time domain correlation with a calculated replica based on the input signal. Next, the grazing angle was determined from the delay between the direct path arrival and reflected path arrival. A temporally isolated direct path signal was subtracted from the recorded time series leaving only the reflected path signal and associated noise. Then, a second correlation was preformed using a temporally isolated direct path signal as the replica. The signal was then bandpass filtered on the data to isolate each frequency band. The initial specular bistatic amplitude and phase were determined by comparing the reflected path amplitude and phase at the peak of the correlation to the direct path amplitude and phase. Last, the data were corrected for transducer response, beam pattern, and spherical spreading by normalizing with the air/water interface response determined in the same manner.

A. Replica correlation and time compression

For this work, the replica correlation and associated time compression were accomplished using a Wiener filter based on the expected signal, x_{DP},

$$F_{RP,DP}(f, \theta_0) = \frac{X^*_{DP}(f, \theta_0)}{|X_{DP}(f, \theta_0)|^2 + \varepsilon^2}X_{RP,DP}(f, \theta_g)\text{BP}(f). \tag{22}$$

Here $F_{RP,DP}$ is the Fourier transform of the filtered signal. X_{DP} is the Fourier transform of the expected signal. $X_{RP,DP}$ is the Fourier transform of the raw data and BP (f) is the pass band. ε is the noise floor.

A sample Wiener filtered time series is shown in Fig. 6. The Wiener filter is constructed so that the time delay of the direct and reflected paths are absolute.

Frequency resolution is established by band pass filtering the direct and reflected paths. The pass bands were spaced every 12 kHz. Chebyshev windows in the frequency domain were used to enhance the time compressed peaks and suppress side lobes in the time domain.

The disadvantage to using the Wiener filter method is the limited frequency resolution. The peak width is inversely proportional to the bandwidth of the signal. If the bandwidth is decreased, then the peak width in the time domain will increase. If the bandwidth is too small, the time series width becomes too large to distinguish the reflected path arrival from scattered arrivals from the walls of the tank. The bandwidth of the band pass filters was optimized so that the direct and reflected peaks were resolvable.

B. Determining the grazing angle

The grazing angle is calculated by taking the arc-cosine of the direct path length over the reflected path length. Since the sound speed in the water column is relatively constant, distance and time are proportional. Therefore, taking the ratio of the direct and reflected path length is equivalent to taking the ratio of the direct and reflected arrival times, t_{DP} and t_{RP}, respectively,

$$\theta_g = \cos^{-1}\left(\frac{t_{DP}}{t_{RP}}\right). \tag{23}$$

Figure 6 shows how Wiener filtering over the entire frequency range (30–160 kHz) was used to precisely locate the arrival times of the direct and reflected paths. The figure also illustrates how the change in height above the bottom increases the grazing angle and how the reflection coefficient (R_x) may be determined by the ratio of the peaks height of the correlated signal. In this figure, the initial reflection coefficient for $23.5°$ is -2 dB and for $45.5°$ it is -11 dB.

For the determination of the grazing angle, the expected signal used in the Wiener filter was the input signal modified by the amplitude of the direct path. Since the DAC and ADC clocks are synchronized, the transmitted input signal starts simultaneously with the actual recorded measurements. This absolute reference was critical to accurately determining the proper arrival time.

C. Determining the reflection coefficient

The initial reflection coefficient $[R(f, \theta_g)]$ is determined by the ratio of the peak of the reflected path correlation and the direct path correlation. This has the added benefit of accounting for transducer response.

The beam pattern and spreading losses were accounted for by normalization with the data taken from the air/water interface. Assuming that the noise is negligible, the ratio of the reflection coefficient measurement from the air/water interface η_a and that of the water/sediment interface η_s is

$$\frac{\eta_s}{\eta_a} = \frac{B(f, \theta_g)}{B(f, \theta_g)} \frac{S_p(\theta_g)}{S_p(\theta_g)} \frac{R_a(f, \theta_g)}{R_s(f, \theta_g)}. \tag{24}$$

Since the complex reflection coefficient for the air/water R_a interface is known, normalizing the sediment initial reflection coefficient by the air/water initial reflection coeffi-
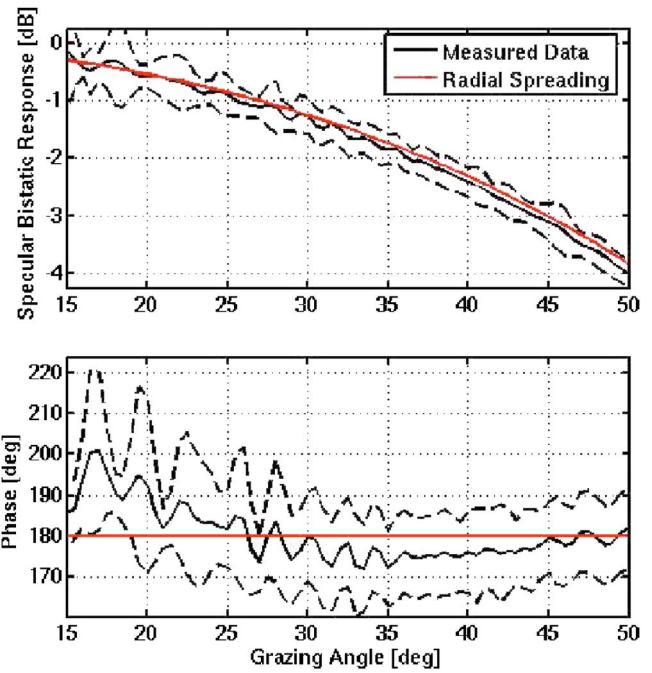
FIG. 7. Magnitude and phase of the 1.27 m air/water interface at 54 kHz, which includes only spreading and beam pattern effects.

cient removes both the frequency and angular dependent beam pattern and the spherical spreading. For normalization, all initial reflection coefficients are first interpolated to a standard angle and frequency set. Then both the air/water interface and sediment/water interface data are averaged over all spatial locations.

D. Phase determination and normalization

For the phase measurement, the Hilbert transform of the time series was taken prior to filtering. This created a complex time series. The phase angle was determined at the peak of the correlation with the Wiener filter in the time domain. The absolute phase was determined by subtracting the phase of the direct path from the reflected path. As with the magnitude, the phase was corrected for beam pattern effects by normalizing by the air/water interface data.

V. RESULTS

A. Normalization measurements

Ten normalization measurements were taken over the entire angle range at each range by inverting the experimental frame and measuring the specular bistatic response from the air/water interface. These data were averaged and used to verify that the grazing angle calculations were valid and to remove beam pattern effects from the sediment measurements. Figure 7 shows the phase and magnitude for the 1.27 m data for the air/water interface at 54 kHz uncorrected for spreading losses. The data were averaged over nine sets. The dashed lines represent the 95% confidence bounds. The overall trend of the data matches well with theoretical spherical spreading ($1/r$). This indicates that the algorithms used to calculate the grazing angle are reliable. The variations can be attributed to the beam pattern. It is evident that the effects of
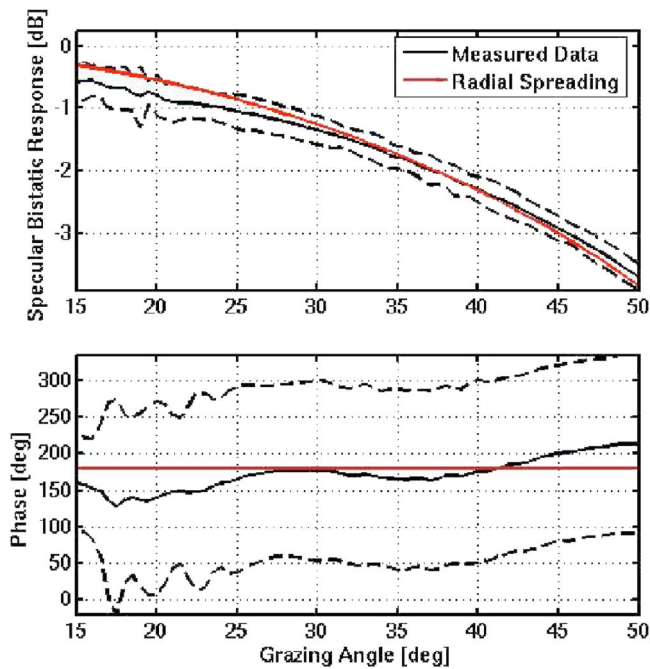
FIG. 8. Magnitude and phase of the 1.27 m air/water interface at 125 kHz compared with theoretical spreading. Deviations are indicative of beam pattern effects.

the beam pattern are small compared to the effects of the spreading. The beam pattern exhibits 0.1 dB oscillations especially in the 25–30 angle range at 54 kHz as evident in Fig. 7. These oscillations are not apparent at larger frequencies as seen in Fig. 8.

The frequency dependent beam pattern of the two phones was calculated by subtracting the theoretical spreading from the measured air/water interface reflection data. For the phase, the beam pattern effects is determined by subtracting 180° from the phase. The 1.27 m data set has slight variations of approximately ±1 dB for a 95% confidence interval over the entire frequency range. The phase difference can be as much as 50° depending on the frequency and angle. The error is generally around 10° when averaged over the broadband. The 0.25 m data set has similar characteristics up to approximately 80 kHz. At the higher frequencies, the 0.25 m data becomes unstable. This is shown in Fig. 9 for 125 kHz. The 0.25 m data may be more heavily influenced by volume properties due to a higher influence of the sediment born lateral wave and steeper ray angles which may penetrate further into the sediment. Therefore, only the 1.27 m data will be considered quantitatively.

B. Spherical wave effects

The normalized reflection coefficient results for the 1.27 m case and the 0.25 m case are shown in Figs. 10 and 11 for 54 kHz band with a 95% confidence interval. These are spatial averages of 30 1.27 m measurements and 35 0.25 m measurements along the sediment floor in the calibration tank. Although the critical angle should be located at approximately 30°, both data sets are heavily influenced by spherical wave effects which shifts the critical angle value. This inward shift of the critical angle, especially at lower
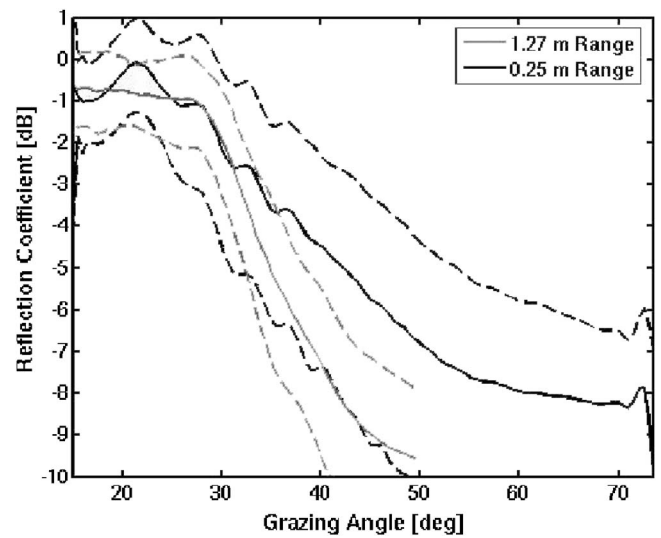
FIG. 9. A comparison of the variability of the 1.27 m range data with the 0.25 m range data at 125 kHz. The spatial instability of the 0.25 m may imply a stronger dependence on volume heterogeneities due to steeper, more penetrating rays and an increased influence of the sediment born lateral wave.

frequencies, as well as the interference pattern in the subcritical region of the 1.27 m data are indicative of lateral wave interference.[10]

To demonstrate these effects, the plane wave elastic reflection coefficient and the spherical wave elastic reflection coefficient are calculated using plane wave decomposition and the visco-elastic model and plotted with the data (see Figs. 10 and 11). It is clear that predicting the data with only the plane wave model would lead to large errors especially in the critical region. The plane wave model overpredicts the location of the critical angle by 5° for the 1.27 m data and 12° for the 0.25 m data. The spherical wave model correctly
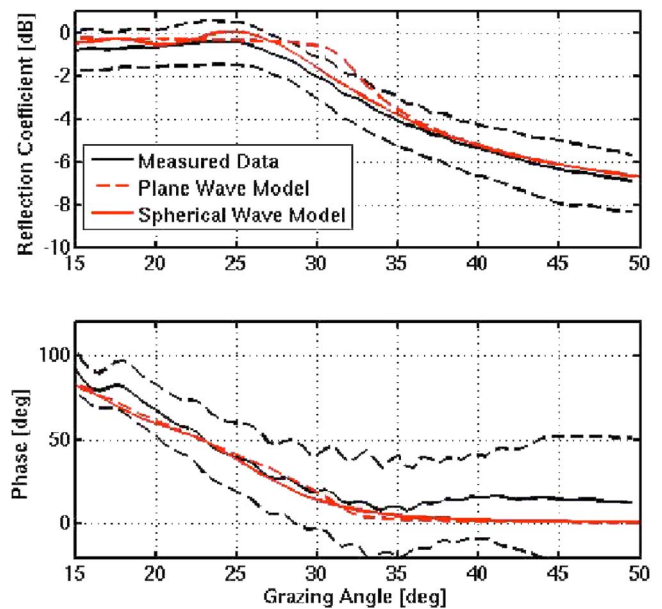
FIG. 10. Plane wave models compared to actual measurements at 1.27 m, 54 kHz. The dashed light line indicates the plane wave model while the solid line indicates the model as calculated using plane wave decomposition. The spherical wave model (solid line) is a better predictor of the data.
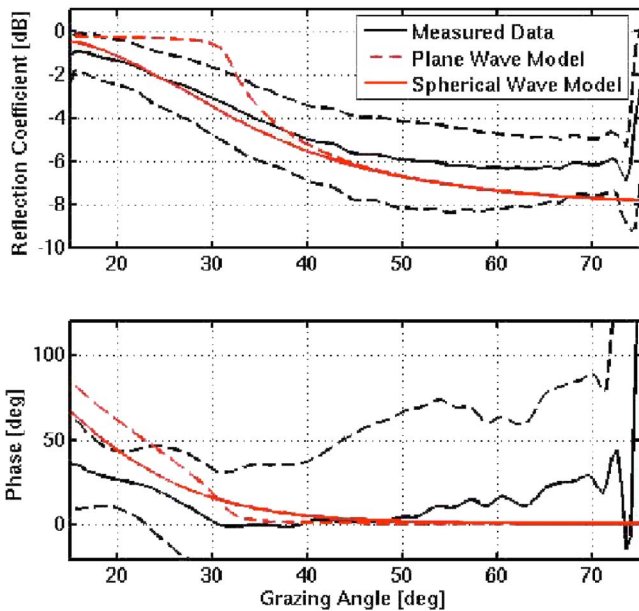
FIG. 11. Plane wave models compared to laboratory measurements at 0.25 m, 54 kHz. The dashed light line indicates the plane wave model while the solid line indicates the model as calculated using plane wave decomposition. The spherical wave model (solid line) is a better predictor of the data.

predicts the decrease in critical angle as well as the presence of an interference pattern subcritically. Additionally, the spherical wave model correctly predicts the dependence of spherical wave effects on transducer spacing although the 0.25 m data is higher than either model as the grazing angle approaches normal. The spherical wave effect is not present at angles larger than critical. The change in phase due to spherical wave effects is small and is not measured in this experiment. For the 1.27 m data both the plane wave model and spherical wave model correctly predict the phase. For the 0.25 m data, the phase is not predicted by either model in the critical region.

In all other model comparisons in this work, only the spherical wave reflection coefficients as calculated using plane wave decomposition will be considered. As stated above for the air/water data, the 0.25 m data were found to be unstable and will not considered quantitatively.

C. Calculation of the models for data comparison

A standard set of parameters were used to calculate all of the reflection coefficients based on the three models described above. The parameters were determined based on sand measurements made in the calibration tank and similar sand collected in the SAX'99 experiment, inversion results of the reflection data collected in the calibration tank, and values found in common reference tables. The attenuation in the water column is assumed to be zero. No attempts were made to adjust the parameters to reflect experimental results.

Table I shows the parameters used in the fluid and viscoelastic models. The fluid density was found in common reference tables. The fluid sound speed, measured as part of this study, was 1476 ± 7 m/s. This value corroborates well with common reference tables. The average wet sediment density was calculated based on porosity and the density of

the sand grains and water.[13] The sediment sound speed in the calibration tank was measured[14] and inversion results[15] verified this result. The value for the compressional attenuation was taken from the SAX'99 measurements since the sediments at that site and that of the calibration tank are similar in composition and size distribution.[16] The shear speed was obtained by inversion results[15] and corroborated with measurements at 312 Hz.[17] The shear attenuation, which will be shown to have little effect on the data, was extrapolated from measurements on unsorted sand at lower frequencies.[18]

Table II shows the poroelastic parameters used in the grain shearing model and the effective density fluid model. The frame shear and bulk moduli are assumed to be zero as described in Sec. II. The porosity was obtained by inversion[15] and is consistent with, values measured in the tank[13] and recorded at SAX'99.[16] The fluid bulk modulus was also determined by inversion[15] and substantiated by common reference tables. Viscosity was determined from common reference tables. The grain density was based on a grain analysis in the tank which revealed that the sand consisted of 85% quartz and 15% feldspar*. Grain bulk modulus was obtained by inversion of this data set[15] and is in the 95% confidence interval measured for similar sediment.[20] The value for tortuosity was taken from the SAX 99 measurements[16] although the same value was determined by inversion.[15] The pore size parameter was calculated using the Kozeny-Carmen equation[19] and the permeability was assumed to be similar to the SAX'99 valued.[16]

Table III shows the empirical values used in Buckingham's model. The frequency at which the sediment compres-

TABLE I. Viscoelastic parameters.

Parameter	Value	Source
Fluid density	0.998 g/cm^3	Table
Fluid sound speed	1481 m/s	Table
Sediment density	1.990 g/cm^3	Calculated
Sediment sound speed	1740 m/s	Measured
Comp. attenuation	0.25 dB/m/kHz	SAX 99 data[a]
Shear speed	80 m/s	Inversion[b]
Shear attenuation	2.5 dB/m/kHz	Estimated from Ref. 18

[a]See Ref. 16.
[b]See Ref. 15.

TABLE II. Poroelastic parameters.

Parameter	Value	Source
Porosity	0.392	Measured[a]
Fluid density	0.998 g/cm^3	Table
Fluid bulk modulus	$2.16E+09$ Pa	Inversion[b]
Viscosity	0.001 kg/ms	Table
Grain density	2.650 g/cm^3	Table
Grain bulk modulus	$3.84E+10$ Pa	Inversion
Tortuosity	1.35	SAX'99 data[c]
Pore size parameter	$4.89E-05$ m	Calculated[d]
Permeability	$2.5E-11$ m2	SAX'99 data

[a]See Ref. 13.
[b]See Ref. 15.
[c]See Ref. 16.
[d]See Ref. 19.

TABLE III. The grain shearing model parameters.

Parameter	Value	Source
Comp. frequency	30 kHz	Measured
Comp. sound speed	1740 m/s	Measured
Comp. attenuation	7.5 dB/m	SAX'99 data[a]
Shear frequency	312 Hz	Measured[b]
Shear sound speed	80 m/s	Measured[b]
Shear attenuation	0.9 dB/m	Measured

[a]See Ref. 16.
[b]See Ref. 17.

sional sound speed in the calibration tank was measured was 30 kHz.[14] The frequency of the shear sound speed measurement is 312 kHz.[17] The remaining values are from Table II. The plane wave reflection coefficient (PWRC) models calculated with the common parameters were modified with PWD to include spherical wave effects and then compared to the measured spherical wave reflection coefficient.

D. Data/model comparison

In order to compare the data with the predictive models, three frequency bands were chosen of the ten measured. These bands were centered at 54, 102, and 125 kHz. These data were indicative of the quality of the entire data set. The model predictions compared with the data are shown in Figs. 12–14. The data are shown with a 95% confidence interval.

Qualitatively, at 54 kHz, (Fig. 12), the elastic and grain shearing models are all underpredicting the bottom loss by almost 1 dB from the mean although these models do fit within the error. The EDFM, however, overpredicts the losses above the critical angle. However, this model is also

FIG. 13. Spherical wave models compared to measurements at 1.27 m, 102 kHz. All models lie within the 95% confidence interval.

within the 95% confidence interval. All three models predict the critical angle within about 3° and correctly predict the phase behavior.

For the higher frequencies, the elastic and grain shearing models under predict the value of the reflection coefficient at high grazing angles while the EDFM model predicts the mean value. In fact, both the fluid and grain shearing models predict values that are not within the 95% confidence interval of the data. EDFM model predicts a lower than measured

FIG. 12. Spherical wave models compared to measurements at 1.27 m, 5 Hz. All models lie within the 95% confidence interval.
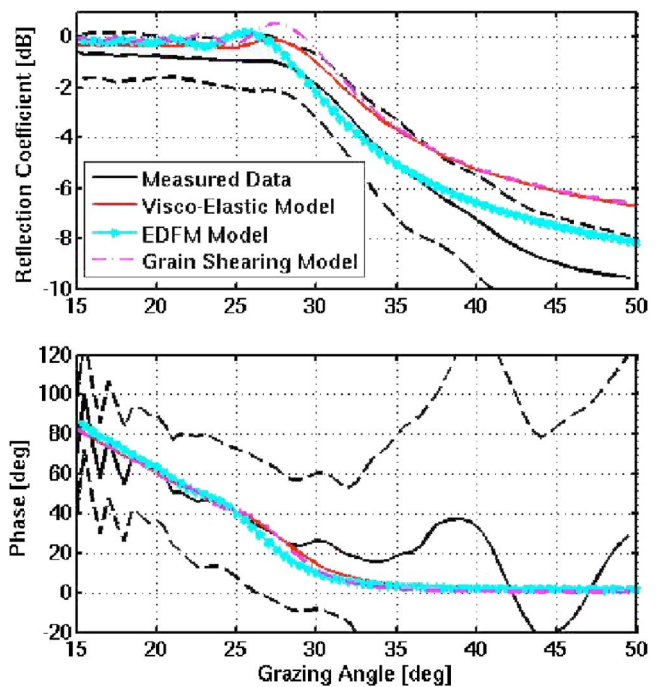
FIG. 14. Spherical wave models compared to measurements at 1.27 m spacing, 125 kHz. Only the poroelastic, EDFM lies within the 95% confidence interval.

TABLE IV. 1.27 m average error results.

Frequency	54 kHz	102 kHz	125 kHz
Elastic (%)	8.1	6.3	10.7
Grain shearing (%)	10.2	8.3	12.8
EDFM (%)	8.1	5.6	5.3
Angle	15°	30°	45°
Elastic (%)	4.6	6.6	15.4
Grain shearing (%)	7.1	11	15.9
EDFM (%)	6.3	7.7	6.8

FIG. 15. Sensitivity analysis of three parameters for the poroelastic model. Shown in the figure are all of the combinations of the parameter ranges from Table V compared to the measured data and the elastic model predictions. In (a) are shown the angle dependent measurements and predictions at 125 kHz. In (b) are shown the frequency dependent predictions at 50 deg. There is still significant difference between the poroelastic model predictions and the elastic model at high frequencies and grazing angles.

value for of the reflection coefficient at the critical angle, however, the prediction is still within the 95% confidence interval. All three models correctly predict the phase behavior at 54 kHz although the phase measurement becomes more unstable at higher frequencies. At 125 kHz none of the models predict the phase behavior subcritically.

The three models were quantitatively compared by taking the mean absolute relative error averaged over the angle range at 54, 102, and 125 kHz. The mean relative error was also considered averaged over frequency at three angles, 15°, 30°, and 45°. The results are shown in Table IV.

The elastic model performs slightly better at low angles, while the poroelastic model is significantly more accurate at higher angles and high frequencies. This is in agreement to the results from Williams who found that at lower frequencies, the water is locked to the frame and the total mass density is in closer agreement with the effective density.[3]

A sensitivity analysis was performed on the EDFM model for three parameters: the fluid bulk modulus, the grain bulk modulus, and the tortuosity. The analysis was performed since the fluid bulk modulus and the grain bulk modulus were determined from an inversion of the same data set and the tortuosity was measured within a range at SAX 99.[16] The range of parameters tested are given in Table V. The fluid bulk modulus range is determined to be a proportional range as determined at SAX 99 for seawater. The range of the grain bulk modulus is taken from the 95% confidence interval of the measurements in Ref. 19. The tortuosity range is given by the measurement at SAX 99.[16] All possible combinations of these parameters were calculated and the range of solutions is shown in Fig. 15 compared with the measured data and the elastic solution for all angles at 125 kHz [Fig. 15(a)] and for all frequencies at an angle of 50 deg. [Fig. 15(b)]. As shown in the figure, the poroelastic model at all range combinations is significantly different

TABLE V. Ranges for the three underdetermined parameters from the poroelastic model. The fluid bulk modulus has the same proportional range as that measured at SAX 99 (see. Ref. 16). The grain bulk modulus range was measured in 1998 by Briggs et al. (see Ref. 20). The range for the tortuosity was measured at SAX 99 (see Ref. 16).

Parameter	Min.	Max.
K_f-Gpa	2.14	2.18
K_r-Gpa	32	49
τ-m^2	1.19	1.57

from the elastic model at high grazing angles and a better fit for the data at high grazing angles and frequencies.

E. Data variance

The standard deviation at high grazing angles and high frequencies in the data is typically 2 dB or more. The variation of the data as measured with the scintillation index is highly dependent on both frequency and angle as shown in Fig. 16.

There are several effects that may contribute to the variation. These include interface scattering, volume scattering, layering, and gradients and variation of bulk sediment properties. Additional sources of error associated with the experimental technique or data analysis are not suspected
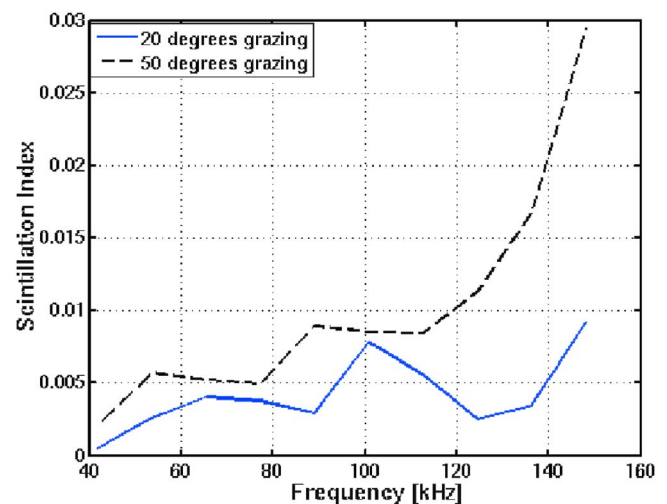
FIG. 16. Scintillation index of the data at 20 and 50 deg grazing as a function of frequency. The variability has a strong dependence on the frequency and grazing angle.
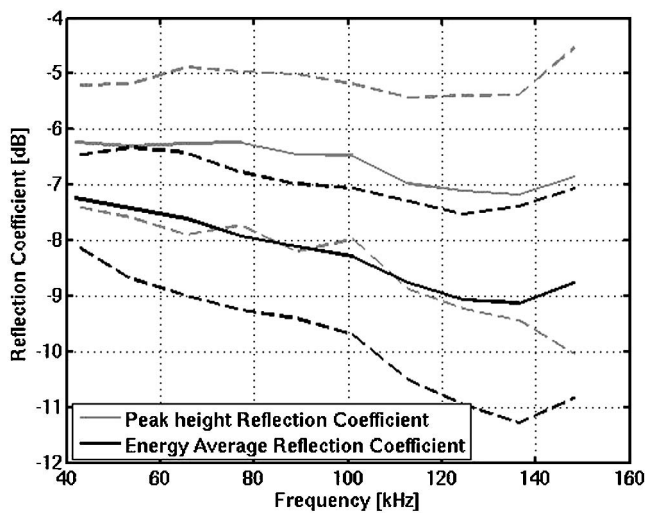
FIG. 17. A comparison of two methods of determining the reflection coefficient: (1) computing the ratio of the reflected path peak height of the Weiner filtered response to that or the direct path and (2) computing the ratio of the reflected path peak area of the Weiner filtered response to that of the direct path. The second method, energy average, is considered the upper bound of the reflection coefficient as it may include effects of volume scattering and multiple layers. The peak height method represents the lower bound of the reflection coefficient.

because of the extremely consistent measurements of the air/water interface. As shown in Fig. 7, the standard deviation of the air/water interface is less than 0.1 dB for the 1.2 m data.

If the interface roughness is small compared to the acoustic wave length, scattering can be considered a stationary random process and the mean value of the reflection coefficient should not be affected.[21] However, for this experiment, the Rayleigh parameter was not much less than 1 for all frequencies. An estimate of the contribution of the interface scattering can be obtained by comparing the peak height measurement of the reflection coefficient with the energy average value. The peak height measurement, which is used throughout this study, is determined by calculating the ratio of the reflected path peak height of the intensity of the replica correlation to that of the direct path. The energy average value is made by computing the ratio of the area under the reflected path peak of the intensity of the replica correlation to that of the direct path. The energy average measurement will include contributions of the scattered energy which will arrive at slightly different times than the energy that is directly reflected. However, the energy average may also include contributions from multiple layers and is considered the upper bound for the value of the reflection coefficient. In contrast, the peak height measurement will only include energy that is reflected along the shortest path and is considered the lower bound for the reflection coefficient. Shown in Fig. 17 are the energy average measurement in light gray compared with the peak height measurement in black at 50 deg grazing. The 95% confidence interval is indicated by the dashed lines. As shown in the figure, the energy average reflection coefficient is about 1–2 dB higher than the peak height reflection coefficient. It is unknown whether this effect is caused by interface scattering, multiple layers, or volume scattering. This issue will be considered in subsequent publications.

Bulk variations as defined by the fluid or viscoelastic models are not suspected to account for the high degree of variation in the data. In fact, if these variations were due to spatially dependent bulk sediment properties, there would be roughly a 230% difference in the local sediment impedances along the tested area. This equates to sound speed and density changes on the order of 1200–3200 m/s and 0.9–5.4 g/cm^3, respectively. These limits are significantly outside the realm of independent measurements.

The analysis of the effects which are causing the high variation in the data is outside the scope of this paper and will be considered in subsequent publications.

VI. CONCLUSION

PWD was used to modify three plane wave reflection coefficient models to include spherical wave effects. The resulting spherical wave reflection coefficients were compared to experimental measurements. The measurements were taken from unwashed river sediment from 5° to 75° grazing at frequencies from 30 to 160 kHz. Two different transducer ranges were used, 1.27 and 0.25 m. Thirty different locations were tested to insure statistical significance.

The data were shown to have significant contributions from spherical wave effects that were dependent on frequency and range. The plane wave models overpredicted the location of the critical angle by 5° and 12° for the 1.27 and 0.25 m data, respectively. The spherical wave model as calculated by plane wave decomposition correctly predicted the location of the critical angle. Plane wave decomposition did not significantly alter the phase prediction.

The 1.27 m data were compared with three current models for acoustic interaction with sandy sediments. Each model was calculated using plane wave decomposition to correct for spherical wave effects. It was found that the viscoelastic model and grain shearing model predictions were not within the 95% confidence interval at high frequencies and angles.

The data exhibited high variance which was frequency and angle dependent. This variance is not likely to be caused by variations in bulk properties as defined by the fluid or viscoelastic models. The cause of the variance will be considered in subsequent publications.

In conclusion, these reflection coefficient measurements suggest that plane wave decomposition is a viable method for computing the effects of spherical waves. Furthermore, the data suggest that the poroelastic model is the best predictor for the reflection coefficient from sandy sediments.

[*]As measured by James Piper, 1992.

[1]L. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980).

[2]M. Buckingham, "Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am., **108**, 1796–2815 (2000).

[3]K. Williams, "An effective density fluid model for acoustic propagation in sediments derived from Biot theory," J. Acoust. Soc. Am., **110**, 2276–2281 (2001).

[4]L. Kinsler, A. Frey, A. Coppens, and J. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982).

[5]M. Biot, "Theory of deformation of a porous viscoelastic anisotropic solid," J. Appl. Phys., **27**, 459–467 (1956).

[6]M. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid I. Low frequency range," J. Acoust. Soc. Am., **28**, 168–178 (1956).

[7]R. Stoll, "Reflection of acoustic waves at a water-sediment interface," J. Acoust. Soc. Am., **70**, 149–157 (1981).

[8]A. Wood, *A Textbook of Sound* (G. Bells and Sons, London, 1911).

[9]F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics*, 2nd ed. (Springer-Verlag, New York, 2000).

[10]E. K. Westwood, "Complex ray methods for acoustic interaction at a fluid-fluid interface," J. Acoust. Soc. Am., **85**, 1872–1884 (2002).

[11]S. Chin-Bing and J. Davis, "Nature of the lateral wave effect on bottom loss measurements," J. Acoust. Soc. Am., **71**, 1433–1437 (1982).

[12]S. Varghese and M. Isakson, "The calibration of a laser light sheet method for determining local interface roughness of the ocean floor," IEEE J. Ocean. Eng., **30**, 463–467 (2005).

[13]T. Muir, C. Horton, and L. Thompson, "The penetration of highly directional acoustic beams into sediments," J. Sound Vib., **64**, 539–551 (1979).

[14]N. Chotiros and M. Isakson, "A broadband model of sandy ocean sediments: Biot–Stoll with contact squirt flow and shear drag," J. Acoust. Soc. Am., **116**, 2011–2022 (2004).

[15]A. Worley, "The impact of experimental error and sediment models on reflection coefficient inversion, Masters thesis," Ph.D. dissertation, University of Texas at Austin, 2004.

[16]K. Williams, D. Jackson, E. Thorsos, D. Tang, and S. Schock, "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media," IEEE J. Ocean. Eng., vol. **27**, pp. 413–428 (2002).

[17]B. Luke, "In situ measurement of stiffness profiles in the seafloor using the spectral-analysis-of-waves (SASW) method," Technical Report under ARL:UT Independent Research and Development Program, Austin, TX, 1995.

[18]B. Brunson, "Shear wave attenuation in unconsolidated laboratory sediments," in *Shear Waves in Marine Sediments*, edited by J. Hovem, M. Richardson, and R. Stoll (Kluwer Academic, Dordecht, 1991), pp. 141–148.

[19]P. Carmen, *Flow of Gases Through Porous Media* (Academic, New York, 1956).

[20]M. Richardson, K. Williams, K. Briggs, and E. Thorsos, "Dynamic measurement of sand grain compressibility at atmospheric pressure: Acoustic applications," IEEE J. Ocean. Eng., **27**, 593–601 (2002).

[21]K. Williams, S. Kargl, E. Thorsos, and D. Tang, "Synthetic aperture sonar (SAS) imaging and acoustic scattering strength measurements during SAX04 (Sediment Acoustics eXperiment—2004): Experimental results and associated modeling," in *Boundary Influences in High Frequency, Shallow Water Acoustics* (University of Bath, Bath, 2005), pp. 127–135.

A fast near-field method for calculations of time-harmonic and transient pressures produced by triangular pistons

Duo Chen, James F. Kelly, and Robert J. McGough[a)]

Department of Electrical and Computer Engineering, Michigan State University,
East Lansing, Michigan 48824

Analytical expressions are demonstrated for fast calculations of time-harmonic and transient near-field pressures generated by triangular pistons. These fast expressions remove singularities from the impulse response, thereby reducing the computation time and the peak numerical error with a general formula that describes the near-field pressure produced by any triangular piston geometry. The time-domain expressions are further accelerated by a time-space decomposition approach that analytically separates the spatial and temporal components of the numerically computed transient pressure. Applied to a Hanning-weighted input pulse, time-space decomposition converts each spatio-temporal integral into six spatial integral evaluations at each field point. Time-harmonic and transient calculations are evaluated for an equilateral triangle with sides equal to four wavelengths, and the resulting errors are compared to pressures obtained with exact and approximate implementations of the impulse response method. The results show that the fast near-field method achieves smaller maximum errors and is consistently faster than the impulse response and methods that approximate the impulse response. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2356839]

I. INTRODUCTION

As increasingly complex transducer geometries are adopted for emerging applications of ultrasound imaging and therapy, new methods are needed for rapid calculations of pressure fields produced by these transducers. Fast numerical calculations are especially important for simulations of phased array structures containing hundreds or thousands of transducers that generate pressures in large computational domains. Simulation methods that either directly evaluate the impulse response[1–4] or subdivide each transducer into smaller subelements and then superpose the pressure[5,6] provide a convenient model for these calculations, but in the near-field region, the convergence of these methods is relatively slow.[7] Furthermore, numerical implementations of the impulse response for flat unfocused transducers encounter some difficulties throughout the paraxial region.[8–10]

The numerical problems with the impulse response are eliminated with the fast near-field method (FNM) defined previously for circular[9] and rectangular[10] sources. The FNM also reduces the numerical error and the computation time relative to calculations that employ the impulse response. Further reductions in the computation time with FNM expressions are achieved through a time-space decomposition approach that separates the temporal and spatial components for transient pressure calculations.[7] The resulting expressions contain a small number of spatial integrals that are evaluated numerically and then multiplied by analytical temporal terms. The time-space decomposition approach applied to a circular piston accelerates transient pressure calculations by exploiting the structure of the FNM integral, achieving the same numerical error as the impulse response in much less time.

The substantial reduction in computation time demonstrated by the FNM for calculations of near-field pressures generated by circular and rectangular pistons motivates the derivation of similar integral expressions for triangular sources. After the impulse response is obtained for right, acute, and obtuse triangular sources, general FNM expressions for time-harmonic and transient inputs are then demonstrated for a triangular source, and the time-space decomposition of the FNM integral is presented for a transient excitation. Based on these expressions for the near-field pressure generated by a triangular source, computation times are evaluated for the same peak numerical errors. For time-harmonic inputs applied to a triangular source, results show that FNM calculations are several times faster than both exact and approximate impulse response calculations, and for pulsed excitations, results demonstrate that FNM calculations performed with time-space decomposition are also much faster than exact and approximate impulse response calculations for triangular piston geometries.

II. TIME-HARMONIC AND TRANSIENT NEAR-FIELD PRESSURE CALCULATIONS FOR TRIANGULAR SOURCES

A. Impulse response calculations for a triangular source

The geometry for a right triangular source with a *right angle* $\angle BCA$ at vertex C is depicted in Fig. 1(a). For this right triangle and the triangles in Figs. 1(b) and 1(c), the impulse response is evaluated at a point directly over the

---

[a)]Electronic mail: mcgough@egr.msu.edu

(a) Triangular source with right angle at $\angle BCA$.

(b) Triangular source with acute $\angle BCA$.

(c) Triangular source with obtuse $\angle BCA$.
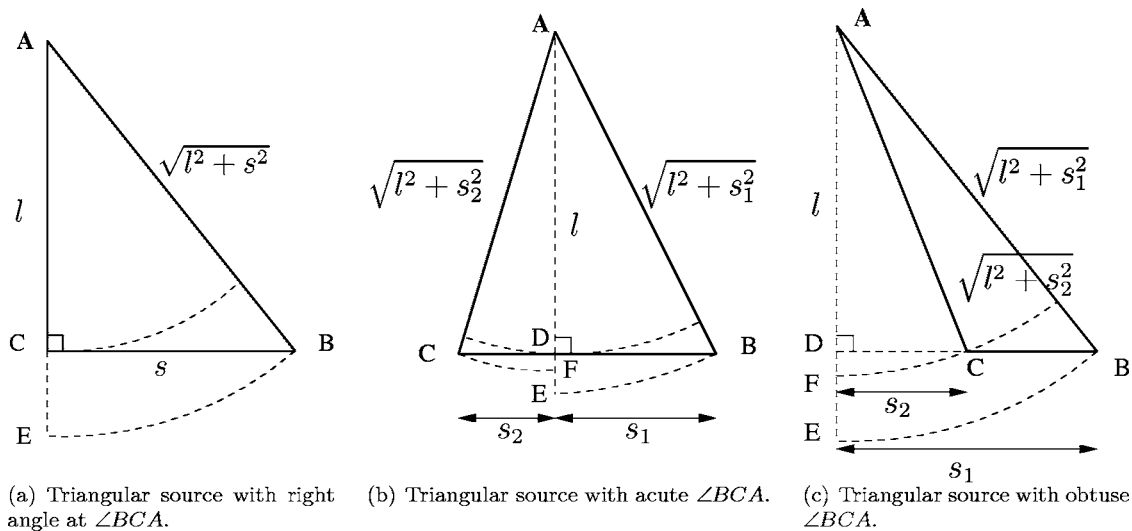
FIG. 1. Triangular source geometries defined for near-field pressure calculations. The near-field pressure is evaluated above the vertex A (indicated in bold), and the shape of the triangle (right, acute, or obtuse) is defined by the angle $\angle BCA$. The height of each triangle is indicated by l, and the bases of the individual right triangles are indicated by s, s_1, and s_2. The acute triangle in (b) is represented by the sum of two right triangles, and the obtuse triangle in (c) is defined as the difference between two right triangles.

vertex A (indicated in bold in Fig. 1), where the orthogonal projection of the observation point onto the source plane is exactly coincident with the vertex A, and the distance from the observation point to the source plane along this orthogonal projection is represented by the variable z. In Fig. 1(a), the acute angle $\angle CAB = \tan^{-1}(s/l)$ defines the angular extent of sector EAB with radius $\sqrt{l^2+s^2}$, which has an impulse response of $c/(2\pi)\tan^{-1}(s/l)$ for $(z/c) \leqslant t \leqslant \sqrt{z^2+l^2+s^2}/c$. The impulse response for the right triangle $\triangle ABC$ contained

within the sector EAB is obtained by subtracting the impulse response of the region ECB between the curved outer edge of the sector and the near edge of the right triangle so that only the contribution from the right triangle $\triangle ABC$ remains. The impulse response of the region ECB is $c/(2\pi)\cos^{-1}(l/\sqrt{c^2t^2-z^2})$ for $\sqrt{z^2+l^2}/c \leqslant t \leqslant \sqrt{z^2+l^2+s^2}/c$, and therefore the impulse response at an orthogonal distance z above the vertex A is

$$
h_{\text{right}}(z;t) = \begin{cases} c/(2\pi)\tan^{-1}(s/l) & \text{for } \dfrac{z}{c} \leqslant t \leqslant \dfrac{\sqrt{z^2+l^2}}{c} \\[2ex] c/(2\pi)\left\{\tan^{-1}(s/l) - \cos^{-1}\dfrac{l}{\sqrt{c^2t^2-z^2}}\right\} & \text{for } \dfrac{\sqrt{z^2+l^2}}{c} \leqslant t \leqslant \dfrac{\sqrt{z^2+l^2+s^2}}{c}. \\[2ex] 0 & \text{otherwise} \end{cases}
\tag{1}
$$

For other triangular sources, the impulse response is readily constructed from the sum or difference between two right triangles. Figure 1(b) contains an example of a triangular source with an *acute angle* $\angle BCA$ at vertex C. The expression for the impulse response evaluated at a point directly over the vertex A is obtained by evaluating the *sum* of the contributions from the right triangles $\angle CDA$ and $\angle BDA$, each with a right angle at vertex D. The resulting impulse response above the vertex A in Fig. 1(b) is represented by

$$
h_{\text{sum}}(z;t) = \begin{cases} c/(2\pi)\{\tan^{-1}(s_1/l) + \tan^{-1}(s_2/l)\} & \text{for } \dfrac{z}{c} \leqslant t \leqslant \dfrac{\sqrt{z^2+l^2}}{c} \\[2ex] \dfrac{c}{2\pi}\{\tan^{-1}(s_1/l) + \tan^{-1}(s_2/l) - 2\cos^{-1}(l/\sqrt{c^2t^2-z^2})\} & \text{for } \dfrac{\sqrt{z^2+l^2}}{c} \leqslant t \leqslant \dfrac{\sqrt{z^2+l^2+s_2^2}}{c} \\[2ex] c/(2\pi)\{\tan^{-1}(s_1/l) - \cos^{-1}(l/\sqrt{c^2t^2-z^2})\} & \text{for } \dfrac{\sqrt{z^2+l^2+s_2^2}}{c} \leqslant t \leqslant \dfrac{\sqrt{z^2+l^2+s_1^2}}{c}, \\[2ex] 0 & \text{otherwise} \end{cases}
\tag{2}
$$

where the values of s_1 and s_2 are selected such that $s_1 \geqslant s_2$.

Similarly, Fig. 1(c) contains an example of a triangular source with an *obtuse angle* $\angle BCA$ at vertex C, where the impulse response is again evaluated at a point directly over the vertex A, but the impulse response is instead evaluated for the *difference* between two right triangles. The impulse response for the triangle in Fig. 1(c) is

$$
h_{\mathrm{diff}}(z;t) = \begin{cases} c/(2\pi)\{\tan^{-1}(s_1/l) - \tan^{-1}(s_2/l)\} & \text{for } \dfrac{z}{c} \leqslant t \leqslant \dfrac{\sqrt{z^2+l^2+s_2^2}}{c} \\[2ex] c/(2\pi)\{\tan^{-1}(s_1/l) - \cos^{-1}(l/\sqrt{c^2t^2-z^2})\} & \text{for } \dfrac{\sqrt{z^2+l^2+s_2^2}}{c} \leqslant t \leqslant \dfrac{\sqrt{z^2+l^2+s_1^2}}{c}, \\[2ex] 0 & \text{otherwise} \end{cases}
\tag{3}
$$

where the values of s_1 and s_2 are selected such that $s_1 > s_2$.

1. Time-harmonic impulse response calcuations

The time-harmonic pressure generated by these triangular source geometries is proportional to the Fourier transform of the impulse response. Therefore, the formula for the time-harmonic pressure generated by the right triangle in Fig. 1(a) is

$$
P_{\mathrm{right}}(z;k) = \frac{\omega\rho_0 v e^{j\omega t}}{2\pi}\left\{\frac{j}{k}\tan^{-1}\frac{s}{l}e^{-jkz}\right.
$$
$$
-\frac{j}{k}\tan^{-1}\frac{s}{l}e^{-jk\sqrt{z^2+l^2+s^2}}
$$
$$
\left.+\int_{\sqrt{z^2+l^2}}^{\sqrt{z^2+l^2+s^2}} e^{-jk\beta}\cos^{-1}\frac{l}{\sqrt{\beta^2-z^2}}d\beta\right\}.
\tag{4}
$$

The time-harmonic pressures that are produced by the remaining triangles depicted in Figs. 1(b) and 1(c) are obtained by adding and subtracting the contributions of two right triangles, as for calculations of the impulse response in Eqs. (2) and (3), respectively.

2. Transient impulse response calculations

Transient near-field pressures are computed with the impulse response through the convolution

$$
p(z;t) = \rho_0 \dot{v}(t) \otimes h(z;t),
\tag{5}
$$

where the time derivative of the particle velocity $\dot{v}(t)$ is evaluated analytically from the excitation pulse $v(t)$, and the convolution $\otimes$ is evaluated with the fast Fourier transform (FFT). In particular, the discrete Fourier transforms of $\dot{v}(t)$ and $h(z,t)$ are computed with the FFT, the results are multiplied, and the inverse FFT is applied to the product. The forward and inverse FFT routines are computed with the Fastest Fourier Transform in the West (FFTW) library.[11]

3. FIELD II

FIELD II is a software package[12] that computes the impulse response either by superposing far-field contributions from small rectangles or by evaluating expressions similar to Eqs. (1)–(3). With both approaches, FIELD II modifies the impulse response according to the area under the impulse

response curve.[13] This modification allows FIELD II to reduce the temporal sampling for impulse response calculations, which are directly applicable to transient and steady-state near-field pressure computations.

4. Smoothed impulse response

The Fourier transform of Eq. (5) is

$$
P(z;\omega) = j\omega\rho_0 V(\omega)H(z,\omega).
\tag{6}
$$

Normally, the excitation $V(\omega)$ is bandlimited, so the high-frequency components in the Fourier transform $H(z;\omega)$ of the impulse response are negligible. To exploit the bandlimited characteristics of the excitation $v(t)$, the formula for a smoothed impulse response is given by[14]

$$
h_{\mathrm{smooth}}(z;t) = \frac{O(c(t+\Delta t/2)) - O(c(t-\Delta t/2))}{2\pi tc\Delta t},
\tag{7}
$$

where $O(ct)$ is the area that formed by the intersection of the transducer and the sphere with radius ct centered at the observation point, and Δt is the length of the rectangular pulse that smooths the analytical impulse response. The constraint

$$
f_{E\,\max} < \frac{1}{\Delta t}
\tag{8}
$$

insures proper smoothing, where $f_{E\max}$ is the highest-frequency component of the excitation pulse, and $\Delta t = 0.02\ \mu s$ in the simulations that follow. The result obtained from Eq. (7) is then directly applied to calculations of the near-field pressure for time-harmonic and transient inputs.

B. The fast near-field method for a triangular source

Integral expressions that describe the fast near-field method (FNM) for a triangular source excited by a time-harmonic input are obtained by replacing the inverse cosine term with the integral form of the inverse tangent and then exchanging the order of integration in the impulse response expressions for right, acute, and obtuse triangles. After defining a new variable of integration and subtracting the singularity at $z=0$ from each integrand, the resulting FNM expression for a *right* triangle [Fig. 1(a)], the *sum* of two right triangles [Fig. 1(b)] that share a common side of length l, and

the *difference* between two right triangles [Fig. 1(c)] that share a common side of length l is

$$P(z;k) = -\frac{\rho c v e^{j\omega t}}{2\pi} \int_{x_C}^{x_B} \frac{l}{\sigma^2 + l^2}$$
$$\times (e^{-jk\sqrt{\sigma^2 + z^2 + l^2}} - e^{-jkz}) d\sigma, \qquad (9)$$

where l represents the height of the triangle, and x_B and x_C represent the x coordinates of B and C, respectively. In Figs. 1(a)–1(c), the values of (x_B, x_C) are $(s, 0)$, $(s_1, -s_2)$, and (s_1, s_2), respectively. Thus, a single FNM expression represents all three triangle geometries in Fig. 1, whereas the impulse response requires a separate expression for each triangle in Fig. 1.

1. Transient FNM calculations

The inverse Fourier transform of Eq. (9) generates the FNM expression for the transient response. The transient pressure generated by a triangular source above the vertex A is represented by

$$p(z;t) = -\frac{\rho c}{2\pi} \int_{x_C}^{x_B} \frac{l}{l^2 + \sigma^2}$$
$$\times \left[v\left(t - \frac{1}{c}\sqrt{z^2 + \sigma^2 + l^2}\right) - v(t - z/c) \right] d\sigma, \quad (10)$$

where the transient excitation is represented by $v(t)$. By retaining the $v(t-z/c)$ term within the integral and subtracting the singularity, Eq. (10) maintains the rapid rate of convergence achieved for time-harmonic calculations with Eq. (9).

2. Time-space decomposition

Transient FNM computations are accelerated by decoupling the temporal and spatial dependence of Eq. (10). The time-space decomposition approach, demonstrated previously for a circular source,[7] expands the delayed input pulse $v(t-\tau)$ in terms of temporal weighting functions $g_n(t)$ and spatially dependent terms $f_n(\tau)$ that depend only on the coordinates of the observation point and the variable of integration σ through $\tau = (1/c)\sqrt{z^2 + \sigma^2 + l^2}$. The decoupled input pulse is thus represented by

$$v(t-\tau) = \text{rect}\left(\frac{t-\tau}{W}\right) \sum_{n=1}^{N} f_n(\tau) g_n(t), \qquad (11)$$

where the time duration of the pulse is indicated by the parameter W. The decomposed pulse in Eq. (11) is then inserted into Eq. (10), and then time-dependent terms are factored out of the integral. The result consists of N edge wave terms specified by

$$E_n = -\frac{\rho c}{2\pi} l \int_{x_C}^{x_B} \frac{f_n(\tau)}{\sigma^2 + l^2} \text{rect}\left(\frac{t-\tau}{W}\right) d\sigma \qquad (12)$$

and a direct wave term given by

$$D = -\frac{\rho c}{2\pi} v(t - z/c) l \int_{x_C}^{x_B} \frac{1}{\sigma^2 + l^2} d\sigma. \qquad (13)$$

The temporal dependence of the edge wave integrand in Eq. (12) is eliminated when the effect of the rect function is instead shifted to the limits of integration. This operation, which restricts the edge wave contributions by only considering those that have reached the observation point without completely passing the observation point, completely removes all temporal variables from the integrand. As a result, calculations of transient pressure fields are converted into the numerical evaluation and subsequent superposition of N spatial integrals that are weighted by analytical time-dependent terms. Further reduction in the computation time is achieved by storing redundant edge wave calculations from Eq. (12) in the matrix

$$K_n(i,j) = \sum_{m=i}^{j} w_m \frac{f_n(\tau[\sigma_m])}{\sigma_m^2 + l^2}. \qquad (14)$$

In Eq. (14), w_m represents the weights and σ_m represents the abscissas computed for Gauss quadrature, the value of $\tau[\sigma_m]$ is obtained from the relation $\tau[\sigma_m] = (1/c)\sqrt{z^2 + \sigma_m^2 + l^2}$, and the indices i and j indicate the shortest and longest times that correspond to the limits of integration. The values in the $K_n(i,j)$ are initialized within the computation procedure only for the points that are needed, and then the time-space decomposition calculations superpose the numerically computed results of the spatial integrals with analytical time-dependent weighting factors to achieve a significant reduction in computation time for transient pressure calculations in the near-field region.

C. Superposition calculations with impulse response and FNM expressions

At observation points away from the normal that passes through a vertex of the triangular source, impulse response and FNM calculations project the observation point onto the source plane and then superpose the contributions from two or three triangles as in Fig. 2. The contributions from three triangles are either added, as shown in Fig. 2(a) for an observation point within the lateral extent of the source triangle $\triangle ABC$, or added and subtracted as demonstrated in Fig. 2(b) for an observation point outside of the lateral extent of the source triangle. Whether a contribution is added or subtracted depends on the location of the projected observation point in the source plane relative to each side of the triangular source.

The FNM admits some additional simplifications for near-field calculations of pressures generated by the triangle $\triangle ABC$ in Fig. 2. If the three lines that are coincident with the three sides of the source triangle $\triangle ABC$ are defined in the general form $E_i x + F_i y + G_i = 0$, then the distances from the projected observation point to each of the three sides are represented by $l_i = |E_i x + F_i y + G_i| / \sqrt{E_i^2 + F_i^2}$. Likewise, the sign of each contribution is defined as $S_i = (E_i x + F_i y + G_i) / |E_i x + F_i y + G_i|$ for coefficients E_i, F_i, and G_i chosen such that S_i is positive within the lateral extent of the source $\triangle ABC$. Furthermore, the lower and upper limits of integra-

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Chen *et al.*: Fast near-field method for triangular pistons   2453
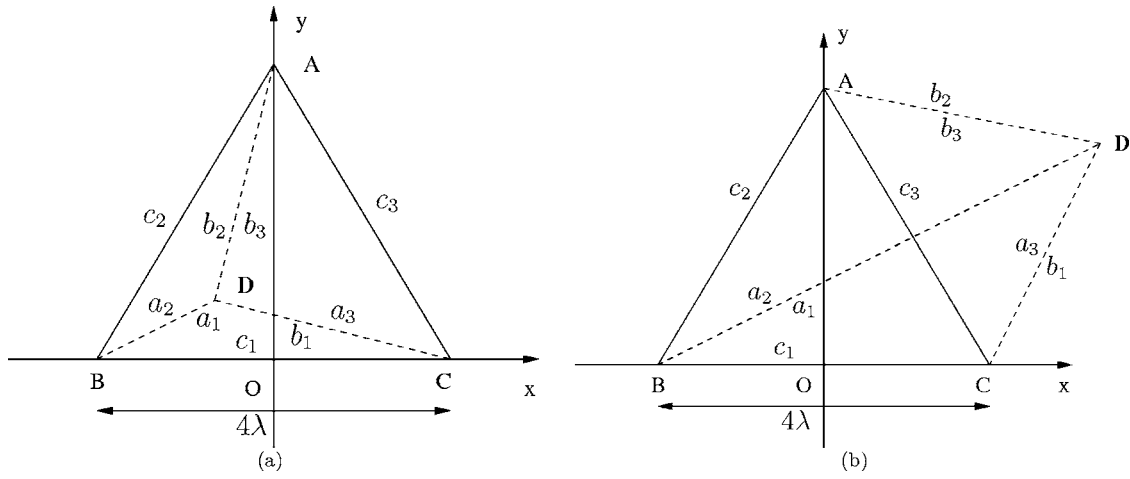
FIG. 2. Superposition operations that calculate near-field pressures generated by the equilateral triangular source ABC, where each side is 4 wavelengths long. The vertex D (indicated in bold) is the projection of the observation point onto the source plane, which partitions the radiating source into three triangles with sides (a_i, b_i, c_i). (a) The field point is located inside of the equilateral triangular source, and the total field is obtained by adding the contributions from the three triangles that share a vertex at D. (b) The field point is located outside of the equilateral triangular source, and the total pressure is obtained by adding and subtracting the contributions from the three triangles that share a vertex at D.

tion are defined as $(a_i^2 - b_i^2 - c_i^2)/(2c_i)$ and $(a_i^2 + c_i^2 - b_i^2)/(2c_i)$, respectively. The resulting near-field pressure generated by $\triangle ABC$ in Fig. 2 is therefore represented by

$$
P(x,y,z;k) = -\frac{\rho c v e^{j\omega t}}{2\pi} \sum_{i=1}^{3} \left(\frac{E_i x + F_i y + G_i}{\sqrt{E_i^2 + F_i^2}} \right.
$$
$$
\left. \times \int_{\frac{a_i^2 - b_i^2 - c_i^2}{2c_i}}^{\frac{a_i^2 + c_i^2 - b_i^2}{2c_i}} \frac{e^{-jk\sqrt{\sigma^2 + z^2 + l_i^2}} - e^{-jkz}}{\sigma^2 + l_i^2} d\sigma \right). \quad (15)
$$

Calculations with Eq. (15) compute the values of c_i, E_i, F_i, and G_i only once for each edge of $\triangle ABC$, whereas the values of a_i and b_i are calculated once for each (x,y) pair. Unlike the expressions for the impulse response that change depending on the spatial coordinate, Eq. (15) is a general formula that computes the near-field pressure with a single expression that is valid at all points in space.

D. Transient input waveform

Evaluations of the impulse response and the FNM with time-space decomposition are performed for the Hanning-weighted pulse specified by

$$
v(t) = \frac{1}{2}[1 - \cos(2\pi t/W)]\sin(2\pi f_0 t)\mathrm{rect}(t/W), \quad (16)
$$

where $\mathrm{rect}(t) = 1$ if $t \in [0,1]$ and $\mathrm{rect}(t) = 0$ otherwise. In the simulations that follow, the input is a Hanning-weighted pulse with a center frequency $f_0 = 2$ MHz and a pulse duration $W = 1.5$ μs. Time-space decomposition performed on this pulse with $N = 6$ yields the entries in Table I, where the spatial edge wave integral in Eq. (12) is evaluated once for each row entry applied to each edge of the source triangle $\triangle ABC$ in Fig. 2, and then the results are weighted by the temporal basis functions in Table I.

E. Error calculations

For time-harmonic near-field pressure calculations, the numerical error $\eta(x,y,z)$ is defined as the normalized difference between the reference field and the computed field according to

$$
\eta(x,y,z) = \frac{|P(x,y,z) - P_{\mathrm{ref}}(x,y,z)|}{\max|P_{\mathrm{ref}}(x,y,z)|}, \quad (17)
$$

where $P_{\mathrm{ref}}(x,y,z)$ is the reference time-harmonic near-field pressure. For transient near-field pressure calculations, the numerical error $\eta(x,y,z)$ between the computed transient field and the reference transient field is defined by

$$
\eta(x,y,z) = \frac{\|p(x,y,z;t) - p_{\mathrm{ref}}(x,y,z;t)\|}{\max_{x,y,z}\|p_{\mathrm{ref}}(x,y,z;t)\|}, \quad (18)
$$

where $\|\cdot\|$ denotes the energy norm used with respect to time, and $p_{\mathrm{ref}}(x,y,z;t)$ is the reference transient pressure field as a function of time. The maximum error is defined as $\eta_{\max} = \max_{x,y,z}\eta(x,y,z)$, and this value is computed for both time-harmonic and transient excitations.

III. RESULTS

All simulation programs are written in C, then compiled and executed within a MATLAB-C language MEX interface. The simulations are performed on an eMachines T3958 per-

TABLE I. Basis functions for time-space decomposition with a Hanning-weighted pulse.

Temporal basis functions $g_n(t)$	Spatial basis functions $f_n(\tau)$
$g_1(t) = 1/2 \sin(2\pi f_0 t)$	$f_1(\tau) = \cos(2\pi f_0 \tau)$
$g_2(t) = -1/2 \cos(2\pi f_0 t)$	$f_2(\tau) = \sin(2\pi f_0 \tau)$
$g_3(t) = -1/2 \cos(2\pi t/W)\sin(2\pi f_0 t)$	$f_3(\tau) = \cos(2\pi \tau/W)\cos(2\pi f_0 \tau)$
$g_4(t) = 1/2 \cos(2\pi t/W)\cos(2\pi f_0 t)$	$f_4(\tau) = \cos(2\pi \tau/W)\sin(2\pi f_0 \tau)$
$g_5(t) = 1/2 \sin(2\pi t/W)\sin(2\pi f_0 t)$	$f_5(\tau) = \sin(2\pi \tau/W)\cos(2\pi f_0 \tau)$
$g_6(t) = 1/2 \sin(2\pi t/W)\cos(2\pi f_0 t)$	$f_6(\tau) = \sin(2\pi \tau/W)\sin(2\pi f_0 \tau)$

Chen *et al.*: Fast near-field method for triangular pistons
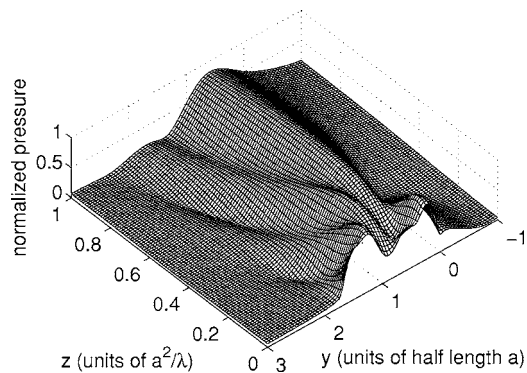
FIG. 3. Simulated time-harmonic pressure field in the $x=0$ plane for an equilateral triangular source with sides equal to 4 wavelengths. The reference field is generated by the impulse response method computed with 100 000-point Gauss quadrature.

FIG. 4. Peak normalized error for calculations of near-field pressures generated by the triangular source in Fig. 2 plotted as a function of the computation time. The results show that the FNM consistently achieves smaller errors in less time than exact and approximate impulse response calculations for time-harmonic excitations.

sonal computer with a 2.93-MHz Celeron D processor. The operating system on this computer is Fedora Core 3 Linux. All simulations are run sequentially under similar operating conditions.

A. Time-harmonic near-field pressure calculations

1. Reference pressure distribution

The reference pressure field is computed in Fig. 3 for an equilateral triangular source with sides equal to 4 wavelengths. In Fig. 3, the acoustic field is evaluated in the $x=0$ plane defined in Fig. 2. The reference near-field pressure distributions in Fig. 3 are obtained when the impulse response is calculated for all triangles with 100 000-point Gauss quadrature. This pressure distribution is selected as the reference because near-field pressures computed with 100 000 abscissas produce normalized errors that converge to 15 significant digits throughout the near-field region, which represents the smallest error achievable with double-precision arithmetic.

2. FNM and impulse response calculations

The numerical errors and computation times for the fast near-field method and the impulse response method are shown in Fig. 4. For the FNM, the exact impulse response, and FIELD II with "use_triangles," near-field pressures are evaluated in an 81×101-point grid in the $x=0$ plane as shown in Fig. 3. FIELD II with "use_rectangles" and the

smoothed impulse response require an offset due to a singularity on the piston face and are therefore evaluated on a smaller 81×86-point grid. The FNM and the exact impulse response are evaluated with Gauss quadrature, and all three integrals corresponding to the three sides of the source triangles are evaluated with the same number of abscissas. The remaining methods that approximate the uniformly sampled impulse response (i.e., FIELD II and the smoothed impulse response) are evaluated with the midpoint rule as described in the user's guide on the FIELD II website (http://www.es.oersted.dtu.dk/staff/jaj/field/). Figure 4 shows that the error for a given computation time is consistently smaller with the FNM, where smaller errors are located nearer to the horizontal axis on the bottom of this log-log plot. Likewise, the time required to achieve a given error is consistently smaller with the FNM, since the FNM plot is consistently located to the left of the impulse response plot.

Comparisons between the impulse response and the FNM evaluated for the same peak error are summarized in Table II. For a 10% peak error, the FNM is 4.39 times faster than the impulse response, and for a 1% peak error, the FNM is 3.44 times faster than the impulse response for this grid and piston geometry. Even greater improvements are observed for smaller peak error values due to the rapid convergence of the FNM. Although these values change somewhat for different source and grid geometries, the FNM is consis-

TABLE II. Number of Gauss abscissas, computation times, and time ratios that describe the reduction in the computation time achieved with the fast near-field method relative to the impulse response and methods that approximate the impulse response for peak errors of 10% and 1%. The FNM and exact impulse response results are evaluated for time-harmonic calculations on a 81×101-point grid located in the $x=0$ plane, and the FIELD II and smoothed impulse response results are evaluated on an 81×86-point grid in the $x=0$ plane that is slightly offset from the transducer face.

	Time-harmonic Near-field computations							
	10% peak error				1% peak error			
	FNM	Impulse response	FIELD II use_triangles	Smoothed impulse response	FNM	Impulse response	FIELD II use_triangles	Smoothed impulse response
Parameters	$N=8$	$N=11$	$f_s=16$ MHz	$f_s=32$ MHz	$N=11$	$N=12$	$f_s=32$ MHz	$f_s=128$ MHz
Time	0.0938 s	0.4112 s	2.7212 s	1.0051 s	0.1286 s	0.4419 s	3.3993 s	4.0130 s
Time ratio	1	4.39	29.02	10.72	1	3.44	26.43	31.21

tently faster than the exact and the approximate impulse response for near-field calculations of time-harmonic pressures.

3. FIELD II calculations

The FIELD II simulation program[12] includes the "use _triangles'" option for calculations that model rectangular and triangular pistons as the superposition of triangular sources. For calculations of time-harmonic pressures with the "use_triangles" option applied to the source geometry in Fig. 2, FIELD II requires a temporal sampling frequency of $f_s = 16$ MHz to achieve a peak error of 10%. The computation time for FIELD II with "use_triangles" is 29.02 times slower than the FNM evaluated on the same grid. For time-harmonic calculations, FIELD II with the "use_triangles" option requires a temporal sampling frequency of $f_s = 32$ MHz to achieve a peak error of 1%. This results in a computation time that is 26.43 times longer than that required for the FNM evaluated on the same grid.

FIELD II also provides a "use_rectangles" option that introduces a numerical singularity on the piston surface, so the pressure is evaluated on a smaller 81×86-point spatial grid that is offset from the piston face. FIELD II with "use _rectangles" evaluated on this reduced grid produces a 10% peak error in 0.8908 s, which is 9.5 times slower than the FNM on the full 81×101-point grid. For a 1% peak error, FIELD II with "use_rectangles" computes the result on the restricted grid in 218.2492 s, which is 1697 times slower than the FNM on the full grid. Only the four fastest methods (on average) are tabulated, so the FIELD II "use_rectangles" calculation results are plotted in Fig. 4 but not included in Table II.

4. Smoothed impulse response calculations

Time-harmonic calculations with the smoothed impulse response[14] evaluate the pressure on a smaller 81×86-point spatial grid that is offset from the piston face. The offset is required for smoothed impulse response calculations so that the singularity in Eq. (7) on the piston face is avoided. For calculations of the time-harmonic pressure generated by the triangular source depicted in Fig. 2 and evaluated within an 81×86-point subset of the grid shown in Fig. 3, the smoothed impulse calculation converges to a peak error of 10% with a temporal sampling rate of $f_s = 32$ MHz. This computation is completed in 1.0105 s, which is 10.74 times longer than the time required for the corresponding FNM calculation evaluated on a larger 81×101-point spatial grid. Time-harmonic calculations with the smoothed impulse response achieve a peak error of 1% for a temporal sampling rate of $f_s = 128$ MHz. This computation is completed in 3.9983 s, which is 31.26 times longer than time required to obtain the FNM result with 1% peak error in Table II.

B. Transient near-field pressure calculations

1. Reference pressure distribution

The reference near-field pressure distribution for transient excitations is calculated with impulse response waveforms that are sampled at $f_s = 524.288$ GHz, zero padded, and
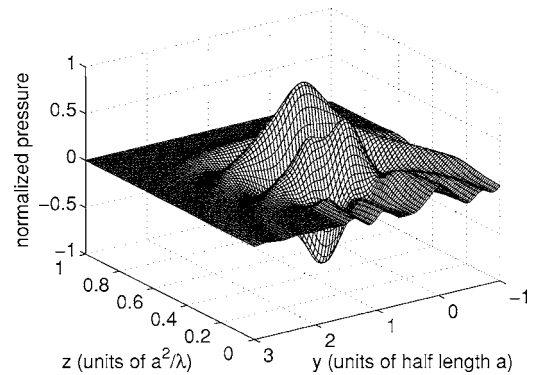
FIG. 5. Simulated transient pressure field in the $x = 0$ plane for a equilateral triangular source with sides equal to 4 wavelengths. For this calculation, the excitation is the Hanning-weighted pulse in Eq. (16), and the transient pressure is evaluated at 85 time points in an 81×101-point grid. The result is plotted at 1.8125 μs after the initiation of the input pulse.

convolved with FFTs. The resulting temporal variations in the near-field pressure, which are evaluated for an equilateral triangular piston with 4 wavelengths on each side, are then downsampled and stored at $f_s = 16$ MHz. The reference field is calculated for a sound speed of $c = 1.5$ mm/μs on an 81 $\times 101$-point spatial grid evaluated at 85 time points, and the result at time $t = 1.8125$ μs is shown in Fig. 5. This error reference is accurate to five significant digits for calculations in the $x = 0$ plane.

2. FNM and impulse response calculations

Figure 6 plots the numerical error as a function of the computation time for the FNM with time-space decomposition and calculations based on the impulse response method. All methods, except for FIELD II with "use_rectangles" and the smoothed impulse response, are evaluated relative to an 81×101-spatial point reference transient pressure distribution computed at 85 time points. FIELD II with "use _rectangles" and the smoothed impulse response are singular at the piston face, so a smaller 81×86-point spatial grid that incorporates an offset from the piston face is again required for transient field computations. The input for the reference
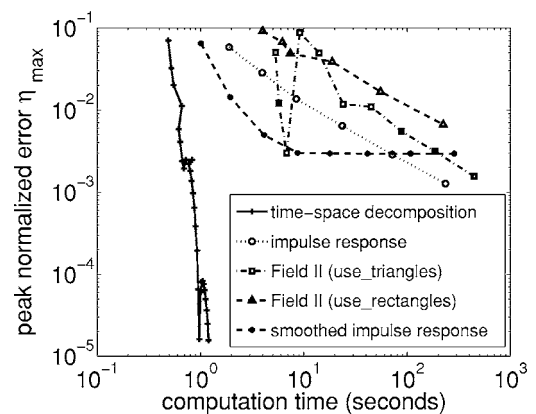
FIG. 6. The peak normalized error plotted as a function of the computation time for the FNM/time-space decomposition method, the impulse response method, and methods that approximate the impulse response. These errors and times are evaluated for transient near-field calculations of an equilateral triangular source with sides equal to 4 wavelengths. The excitation for these calculations is a Hanning-weighted pulse with a center frequency of 2 MHz.

TABLE III. Comparisons of computation times, input parameters, and time ratios that describe the reduction in the computation time achieved with the FNM and time-space decomposition relative to the exact and approximate impulse response for specified maximum errors of 10% and 1%. For FNM, impulse response, and FIELD II calculations with "use_triangles," these transient results are evaluated in an 81×101-spatial point computed at 85 time points, and for the smoothed impulse response, the results are valued at the same temporal points in a restricted 81×86-point spatial grid.

	Transient near-field computations							
	10% peak error				1% peak error			
	FNM	Impulse response	FIELD II use_triangles	Smoothed impulse response	FNM	Impulse response	FIELD II use_triangles	Smoothed impulse response
Parameters	$N=5$	$f_s=128$ MHz	$f_s=16$ MHz	$f_s=32$ MHz	$N=9$	$f_s=1$ GHz	$f_s=64$ MHz	$f_s=128$ MHz
Time	0.4867 s	1.8911 s	5.3317 s	1.0122 s	0.6160 s	23.5241 s	6.8078 s	4.1261 s
Time ratio	1	3.89	10.95	2.08	1	38.19	11.05	6.70

is generated by a Hanning-weighted pulse with a center frequency of $f_0=2$ MHz. The transient near-field pressures are compared for $f_s=16$ MHz, which is the original sampling rate for the FNM calculations and the resulting rate after downsampling for impulse response calculations. Figure 6 shows that the FNM with time-space decomposition is consistently faster than the impulse response and the methods that approximate the impulse response. Similarly, Fig. 6 indicates that the FNM with time-space decomposition achieves much smaller numerical errors than the impulse response and approximations to the impulse response.

Table III shows that the FNM with time-space decomposition achieves a 10% peak error with five Gauss abscissas in 0.4867 s. To achieve a 1% peak error, the FNM with time-space decomposition needs nine Gauss abscissas and the computation time is 0.6160 s. In contrast, the impulse response method achieves a peak error of 10% with a sampling frequency of $f_s=128$ MHz in 1.8911 s. To achieve a peak error of 1%, the impulse response method requires a sampling frequency of $f_s=1$ GHz and a computation time of 23.5241 s. Thus, the reduction in the computation time with time-space decomposition applied to the FNM relative to the impulse response is a factor of 3.89 for a peak error of 10% and a factor of 38.19 for a peak error of 1%.

3. FIELD II calculations

The FIELD II result obtained with the "use_triangles" option for the transient excitation in Eq. (16) requires a sampling frequency of $f_s=16$ MHz to achieve a peak error of 10%, and the computation time for this combination of parameters is 5.3317 s. For a peak error of 1%, FIELD II with the "use_triangles" option requires a sampling frequency of $f_s=64$ MHz, and the computation time is 6.8078 s. Therefore, the FNM with time-space decomposition is 10.95 times faster than FIELD II with "use_triangles" for a peak error of 10% and 11.05 times faster for a peak error of 1%.

Transient FIELD II calculations that subdivide the aperture into small rectangular sources with "use_rectangles" reach a peak error of 10% with a temporal sampling frequency of $f_s=32$ MHz in 3.9926 s. FIELD II with "use_rectangles" achieves a peak error of 1% with a temporal sampling frequency of $f_s=48$ MHz in 221.6569 s. Therefore, the FNM with time-space decomposition is 8.2 times faster than FIELD II evaluated with subdivided rectangular sources for a 10% peak error and 359.81 times faster than FIELD II

evaluated with subdivided rectangular sources for a 1% peak error. As for the time-harmonic results shown in Table II, only the four fastest methods (on average) are tabulated, so FIELD II results obtained with "use_rectangles" are not included in Table III.

4. Smoothed impulse response calculations

Transient calculations with the smoothed impulse response in Eq. (7) evaluate the pressure at 85 time points on a smaller 81×86-point spatial grid that is offset from the piston face. The offset is required in order to avoid the singularity in Eq. (7) on the piston face. For calculations of the time-harmonic pressure generated by the triangular source depicted in Fig. 2 and evaluated within an 81×86-point subset of the grid shown in Fig. 3, the smooth impulse calculation converges to a peak error of 10% with a temporal sampling rate of $f_s=32$ MHz. This computation is completed in 1.0122 s, which is 2.08 times longer than the time required for the corresponding FNM calculation evaluated on a larger 81×101-point spatial grid. Time-harmonic calculations with the smoothed impulse response achieve a peak error of 1% for a temporal sampling rate of $f_s=128$ MHz. This computation is completed in 4.1261 s, which is 6.7 times longer than time required to obtain the FNM result with 1% peak error in Table II.

IV. DISCUSSION

A. Time and error calculations

While computer processor speed and memory has increased substantially in recent decades, the size and complexity of ultrasound therapy and imaging simulations has grown accordingly. Simulations of large ultrasound therapy arrays are now applied to thousands of transducer elements and computational volumes spanning hundreds of wavelengths in three dimensions, and simulations of diagnostic imaging arrays have demonstrated a corresponding increase in the number of active elements and the number of scatterers. As a result, large simulations of ultrasound phased arrays can require 24 h or longer on modern computers.

For these large simulations, the evaluation of computational time and numerical error is essential. The computation time remains the primary bottleneck in these time-consuming calculations, but fair comparisons of computation time also require calculations of the numerical error. In recent years,

evaluations of the numerical error have been neglected due to the slow convergence of the impulse response and methods that approximate the impulse response. Figures 4 and 6 demonstrate this slow convergence, which is further emphasized by the time-harmonic reference field that requires 100 000 Gauss abscissas for convergence to 15 significant digits and by the transient reference field that requires a sampling frequency of $f_s = 524.288$ GHz for convergence to five significant digits.

The rapid convergence of the FNM demonstrated in Figs. 4 and 6 suggests that the FNM is ideal for calculating near-field pressure reference fields. In Fig. 4, time-harmonic FNM calculations converge within 15 significant digits in less than one-third of the time that the impulse response requires for convergence within five significant digits. Likewise, in Fig. 6, transient FNM calculations with time-space decomposition converges within five significant digits in less than one-fifth of the time that the impulse response requires for convergence within two significant digits. In these simulations of a triangular piston source excited by a pulse with a center frequency of 2 MHz, impulse response calculations require a sampling rate of 1 GHz to achieve only two significant digits of accuracy, whereas the FNM with time-space decomposition requires only nine Gauss abscissas applied to each integral and a sampling rate of 16 MHz to achieve two significant digits of accuracy throughout the near-field region.

B. Advantages of the FNM for time-harmonic and transient calculations

The computational advantages of the FNM are obtained from several sources. First, the FNM replaces time-consuming calculations of inverse trigonometric functions with a ratio of polynomials in the integrand. This reduces the computation time without increasing the numerical error. Second, the FNM reduces the numerical error by subtracting a singularity in the integrand. This step, which reduces the numerical error without significantly increasing the computation time, is particularly effective in eliminating numerical problems that occur along the edge of the source and throughout the paraxial region. Third, the FNM defines a single analytical expression that describes the pressure throughout the near-field region, whereas the impulse response requires multiple expressions to define the field generated by a single source. Thus, relative to calculations that employ exact or approximate calculations of the impulse response, convergence is faster with the FNM, and the FNM expressions are easier to evaluate.

The advantage of the FNM with time-space decomposition is that an integral expression with temporal and spatial dependencies is replaced with an equivalent expression that instead evaluates N spatial integrals for each edge of the triangular source and weights the result of each integral with an analytical temporal term. This results in greatly reduced overhead for transient near-field calculations, considering that the impulse response requires sampling rates of $f_s = 128$ MHz for a peak error of 10% and $f_s = 1$ GHz for a peak error of 1% for the source geometry in Fig. 2. The FNM

eliminates these high sampling rates, which therefore facilitates much more efficient utilization of computer memory.

C. FIELD II

The FIELD II calculations with "use_triangles" are evaluated within the same 81×101 spatial grid defined previously for these near-field calculations, whereas the same calculations with subdivided rectangular subapertures (i.e., "use_rectangles") are evaluated in an 81×86 spatial grid that includes an offset from the piston face. The offset is required for these near-field calculations; otherwise, the error grows excessively large on the piston face, which translates into much longer computation times for 10% and 1% peak errors. This occurs because subdividing the aperture introduces a numerical singularity on the piston face. Although FIELD II reduces the sampling frequency relative to other impulse response calculations, the exact impulse response consistently outperforms FIELD II for these time-harmonic near-field calculations, and the FNM evaluated with Gauss quadrature outperforms both of these by a wide margin. Furthermore, the FNM with time-space decomposition is also considerably faster than FIELD II for transient near-field calculations, and the FNM with time-space decomposition, unlike FIELD II with "use_rectangles," allows the computational grid to extend up to the piston face.

D. Smoothed impulse response

Unlike the FNM and the exact impulse response, the smoothed impulse response requires an offset from the piston face for near-field calculations. This offset is required because the denominator in Eq. (7) produces a numerical singularity on the piston face. Despite evaluating the near-field pressure on a smaller 81×86-point spatial grid, the smoothed impulse response is slower than the FNM and the exact impulse response for time-harmonic calculations, as demonstrated in Fig. 4 and Table II. The exact impulse response is faster than the smoothed impulse response for these time-harmonic calculations because the exact impulse response is evaluated with Gauss quadrature, and Gauss quadrature generally converges much faster than other numerical integration methods that uniformly sample the integrand. For transient calculations with both exact and approximate impulse response expressions, uniform sampling is required for convolutions with the FFT. In these transient calculations, the smoothed impulse response gains some advantage over the exact impulse response by evaluating the pressure at a smaller number of spatial grid points and by reducing the problems with aliasing at higher frequencies. Nevertheless, as demonstrated in Figs. 4 and 6, the smoothed impulse response converges more slowly than the FNM for time-harmonic and transient near-field calculations.

V. CONCLUSION

A fast near-field method is presented for numerical calculations of the pressure generated by a triangular source. For time-harmonic near-field computations, the FNM expression in Eq. (9) achieves smaller peak errors in less time than the exact impulse response, the smoothed impulse response,

and the FIELD II program. The results show that the FNM is 4.39 times faster than the exact impulse response for a 10% peak error, and the FNM is 3.44 times faster than the exact impulse response for a 1% peak error. The FNM is at least an order of magnitude faster than FIELD II and the smoothed impulse response for time-harmonic calculations compared at 10% and 1% peak error values. In transient near-field computations, the FNM in Eq. (10) combined with time-space decomposition achieves a substantial reduction in the computation time relative to exact and approximate impulse response calculations for a given peak error value. Transient near-field pressures are evaluated with a Hanning-weighted broadband pulse, and the resulting transient calculation is transformed into the superposition of six spatial integrals. The results demonstrate that the FNM with time-space decomposition is 3.89 and 38.19 times faster than the impulse response for peak errors of 10% and 1%, respectively, evaluated on an 81×101 spatial grid at 85 time points. Comparisons between smoothed impulse response results evaluated on the smaller 81×86-point offset spatial grid and the FNM with time-space decomposition evaluated on the larger 81×101-point spatial grid indicate that the FNM with time-space decomposition is 2.08 times faster than the smoothed impulse response for a 10% peak error, and the FNM with time-space decomposition is 6.7 times faster for a 1% peak error. Compared to the FIELD II program, the FNM is at least an order of magnitude faster for 10% and 1% peak error values. The results also suggest that the FNM, which eliminates the numerical problems that are encountered in exact and approximate impulse response calculations, provides a superior reference for near-field pressure calculations evaluated with time-harmonic and transient inputs.

[1] F. Oberhettinger, "On transient solutions of the baffled piston problem," J. Res. Natl. Bur. Stand., Sect. B, **65B**(1), 1–6 (1961).

[2] P. R. Stepanishen, "Transient radiation from pistons in an infinite planar baffle," J. Acoust. Soc. Am. **49**(5), 1629–1638 (1971).

[3] J. C. Lockwood and J. G. Willette, "High-speed method for computing the exact solution for the pressure variations in the nearfield of a baffled piston," J. Acoust. Soc. Am. **53**(3), 735–741 (1973).

[4] J. A. Jensen, "Ultrasound fields from triangular apertures," J. Acoust. Soc. Am. **100**(4), 2049–2056, 1996.

[5] J. Zemanek, "Beam behavior within the nearfield of a vibrating piston," J. Acoust. Soc. Am. **49**(1), 181–191 (1971).

[6] K. B. Ocheltree and L. A. Frizzell, "Sound field calculation for rectangular sources," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**(2), 242–248 (1989).

[7] J. Kelly and R. J. McGough, "A fast time-domain method for calculating the near field pressure generated by a pulsed circular piston," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**(6), 1150–1159 (2006).

[8] J. A. Jensen, "A new calculation procedure for spatial impulse responses in ultrasound," J. Acoust. Soc. Am. **105**(6), 3266–3274 (1999).

[9] R. J. McGough, T. V. Samulski, and J. F. Kelly, "An efficient grid sectoring method for calculations of the nearfield pressure generated by a circular piston," J. Acoust. Soc. Am. **115**(5), l942–1954 (2004).

[10] R. J. McGough, "Rapid calculations of time-harmonic nearfield pressures produced by rectangular pistons," J. Acoust. Soc. Am. **115**(5), 1934–194l (2004).

[11] M. Frigo, and S. G. Johnson, "The design and implementation of FFTW3," Proc. IEEE **93**(2), 216–231 (2005).

[12] J. A. Jensen, "FIELD: A program for simulating ultrasound systems," *Med. Biol. Eng. Comp., 10th Nordic-Baltic Conference on Biomedical Imaging*, 4, Suppl. 1, Part 1(1), 351–353 (1996).

[13] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**(2), 262–267 (1992).

[14] J. D'hooge, J. Nuyts, B. Bijnens, B. De Man, P. Suetens, J. Thoen, M.-C. Herregods, and F. Van de Werf, "The calculation of the transient near and far field of a baffled piston using low sampling frequencies," J. Acoust. Soc. Am. **102**(1), 78–86 (1997).

On the sound field of a circular membrane in free space and an infinite baffle

Tim Mellow[a)]
Nokia UK Ltd., Farnborough, Hants GU14 0NG, United Kingdom

Leo Kärkkäinen
Nokia Research Center, Helsinki, Finland

An enhanced method for calculating the radiation characteristics of a tensioned circular membrane in free space is presented using an analytical solution to the infinite integral in the free-space Green's function in cylindrical coordinates. This enables direct calculation of the surface pressure series coefficients by equating the coefficients of the resulting Bessel series in a set of simultaneous equations. Eliminating both numerical integration and least-squares minimization improves calculation speed and accuracy. An infinite baffle is introduced to provide an indication of what the theoretical limit of the bass performance would be using a very large enclosure. Furthermore, analytical solutions to the pressure field integrals are presented. A force transmission coefficient is introduced, which is the ratio of the total radiation impedance to the motional impedance. The motional, radiation, and diaphragm impedances of the damped membrane are calculated, together with the near- and far-field pressure responses and efficiency. A comparison is made between the on-axis response (without damping) calculated using this method and using a finite element model. It is demonstrated that good correlation between the two calculation methods can be achieved provided the elements are small enough and a sufficiently large model is used at the frequency extremes. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354041]

I. INTRODUCTION

Interest in electrostatic loudspeakers continues due to their superior performance[1] in terms of low distortion and relatively flat frequency response when suitably damped. Recent advances in digital amplifiers enable improvements in efficiency due to the fact that the reactive current drawn by the static capacitance does not produce the losses that occur with an analogue amplifier. If the loudspeaker has no damping or other kinds of resistance, then there are virtually no losses at all. However, damping, in the form of small holes in the electrodes, is usually applied in order to control the vibrational modes of the membrane. Hence it is useful to derive formulas for the motional input impedance of such a loudspeaker, since this enables its efficiency to be calculated.

The most commonly used model for a loudspeaker is the rigid piston, which assumes global loading over the surface. However, this is not necessarily appropriate in the case of a membrane where there is very strong localized coupling with the acoustic load. Because it is so light, most of the moving mass of the membrane is due to the surrounding fluid medium. Also, because it is acoustically transparent, it would be interesting to see how the pressure field might differ from that of a rigid piston which forms an effectively opaque sound source.

Additionally, the analytical solution to the pressure field of a circular membrane provides a useful benchmark for fluid-structure coupled simulation using the finite element modeling (FEM) method. The latter requires the membrane to be simulated in a virtual anechoic chamber filled with air elements. In this study the analytical solution is used as a reference in order to determine how large this space needs to be in order to match the far-field response. In theory, the anechoic space should be at least as large as the near-field or Fraunhofer zone, which is known to extend further from the source at higher frequencies.[2] Without a baffle, the source becomes a dipole, so the near field extends further from the source at low frequencies too, due to the proximity effect, which, in microphone terminology,[3] is sometimes referred to as "bass tip-up."

The electrostatic circular membrane loudspeaker is one of only a few kinds of transducer which can be completely simulated analytically (except turbulent flow effects between the electrodes and other nonlinear effects are not considered here). By contrast, the electrodynamic speaker is notoriously difficult to model rigorously, due to complicated cone geometries and scattering of the rear wave by the magnet and basket structures. Nearly 20 years after it was first published, Streng's method[4,5] for calculating the sound radiation characteristics of a circular stretched membrane in free space is still the best available. The secret to its success is the ingenious use of a trial function for the surface pressure term of the Kirchhoff-Helmholtz boundary integral, which is derived from Bouwkamp's solution[6] to the free space wave equation in oblate spheroidal coordinates. In fact, this is a somewhat special function since it appears to lead to the simplest possible analytical solution, especially regarding the evaluation of the impedances and near-field pressure. Unlike Streng,

---
[a)]Electronic mail: tim.mellow@nokia.com

though, the current paper presents a Bessel series matching that eliminates the need for numerical integration and least mean square minimization in the solution.

Whereas the sound radiation from a membrane in free space is calculated using the *dipole* part of the Kirchhoff-Helmholtz boundary integral formula[7] together with a power series expression for the surface *pressure*, the sound radiation from a membrane in an infinite baffle is calculated using the *monopole* part of the Kirchhoff-Helmholtz boundary integral formula together with a similar power series expression, but this time for the surface *velocity*.

Hence, the derivation for the latter has some similarities with that of a plane wave passing through a hole in an infinite screen.[6,8–10] The sound scattered from the hole can be considered as that radiated from a resilient disc in an infinite baffle if the (constant) driving pressure over its surface is equivalent to that of the incident wave in the absence of any obstacle. However, unlike the hole or resilient disc, the membrane does not exhibit infinite velocity near its perimeter, due to the fact that it is anchored at its rim. Hence, the solution to the wave equation for the surrounding medium has to be coupled to that of the membrane and solved simultaneously. Suzuki and Tichy[11] solved the coupled equations for a circular plate, but their formulation relied upon numerical integration. In a previous analysis of a membrane by the present authors,[12] numerical integration was avoided, but the least mean squares algorithm was used to solve the coupled equations, which caused some numerical difficulties. In this paper, the deflection trial function is expanded using a Bessel series so that the coefficients can be calculated directly without either least squares minimization or numerical integration.

The parameters chosen for the undamped FEM and analytical responses are the same as those used by Streng[4] for a 250-mm-diam membrane in free space in order to allow direct comparison. Otherwise, analytical characteristics are calculated for a 500-mm-diam membrane with and without damping in order to represent a typical full frequency range commercial design.

II. MEMBRANE IN FREE SPACE

A. Boundary conditions

The equations that follow are written in axisymmetric cylindrical coordinates, with w as the radial ordinate and z as the axial ordinate. The electrostatic driving pressure $\tilde{p}_I$ is used as the input in a wave equation for the membrane. The tilde denotes a harmonically time varying quantity where the factor $e^{i\omega t}$ is suppressed. The membrane deflection $\tilde{\eta}(w)$ (in the z direction) is then used as a parameter to couple it to the surrounding loss-free acoustic medium. Hence the membrane and free space wave equations must be solved simultaneously. The membrane, with radius a, lies in the xy plane as shown in Fig. 1 with its center at the origin and the uniform driving pressure $\tilde{p}_I$ is applied to its surface in the z direction. The pressure field on one side of the xy plane is the symmetrical "negative" of that on the other, so that

$$\tilde{p}(w,z) = -\tilde{p}(w,-z). \tag{1}$$
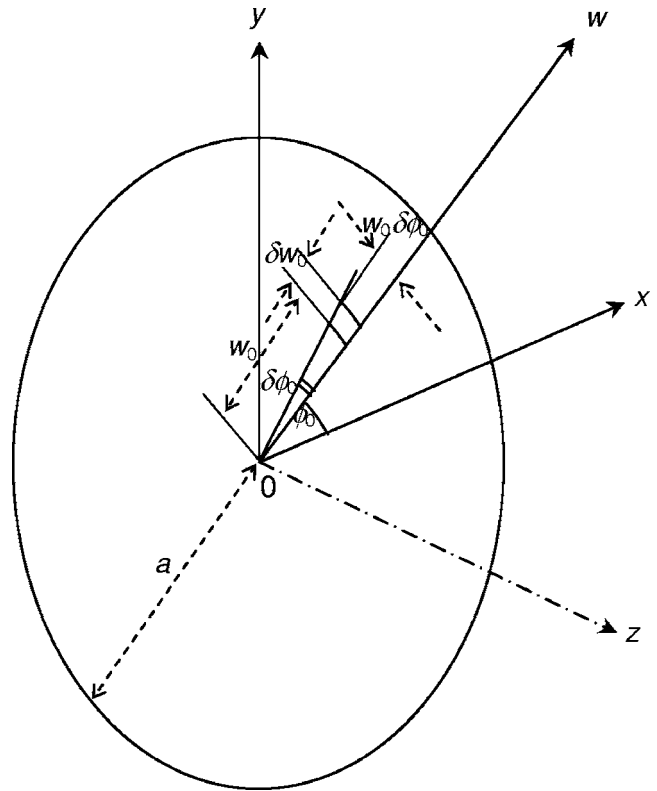
FIG. 1. Geometry of membrane.

Consequently, there is a *Dirichlet* boundary condition in the plane of the membrane where these equal and opposite fields meet, so that

$$\tilde{p}(w,0) = 0, \quad w > a. \tag{2}$$

On the front and rear outer surfaces of the membrane, there is the coupling condition

$$\frac{\partial}{\partial z}\tilde{p}(w,z)\big|_{z=0\pm} = -ik\rho c\tilde{u}_0(w),$$

$$= k^2\rho c^2 \eta(w), \quad 0 \leqslant w \leqslant a \tag{3}$$

where $\tilde{u}_0(w)$ is the normal particle velocity in the z direction at the surfaces and k is the wave number given by

$$k = \frac{\omega}{c} = \frac{2\pi}{\lambda}, \tag{4}$$

where ω is the angular frequency of excitation, ρ is the density of air or any other surrounding medium, c is the speed of sound in that medium, and λ is the wavelength. Values of 1.18 kg/m^3 and 345 m/s are assumed for ρ and c, respectively. The perimeter of the membrane is fixed, which leads to the following boundary condition:

$$\tilde{\eta}(a) = 0. \tag{5}$$

Streng[4,5] showed that the surface pressure distribution $\tilde{p}_+(w_0)$ for any flat axially symmetric unbaffled source (or sink), based upon Bouwkamp's solution[1] to the free space wave equation in oblate spheroidal coordinates could be written as

$$\tilde{p}_+(w_0) = -\tilde{p}_-(w_0) = \sum_{m=0}^{\infty} \tilde{A}_m \left(1 - \frac{w_0^2}{a^2}\right)^{m+(1/2)}, \qquad (6)$$

where $\tilde{A}_m$ are the as yet unknown power series coefficients.

B. Solution to the free space wave equation

1. Rigorous solution

The expression for the pressure field is the same as that given by Eq. (15) (evaluated at $b=a$) in a recent paper on the sound field of disk in a finite circular baffle.[13] The reason why the same formula can be applied to a nonrigid source is that it just describes the sound field due to an arbitrary pressure distribution at the source given by Eq. (6), regardless of the deflection/velocity distribution,

$$\tilde{p}(w,z) = -a \sum_{m=0}^{\infty} \tilde{A}_m \Gamma\left(m + \frac{3}{2}\right) \int_0^{\infty} \left(\frac{2}{\mu a}\right)^{m+(1/2)} J_{m+(3/2)}$$
$$\times (\mu a) J_0(\mu w) e^{-i\sigma z} d\mu, \qquad (7)$$

where

$$\sigma = \begin{cases} \sqrt{k^2 - \mu^2}, & 0 \le \mu < k \\ i\sqrt{\mu^2 - k^2}, & \mu > k \end{cases} \qquad (8)$$

and the power series coefficients $\tilde{A}_m$ are related to normalized dimensionless coefficients τ_m by

$$\tilde{A}_m = \tau_m (m + 3/2) \frac{\tilde{p}_I}{2}. \qquad (9)$$

Applying the boundary condition of Eq. (3) and truncating the power series limit leads to an expression for the surface particle displacement $\tilde{\eta}(w)$ as follows:

$$\tilde{\eta}(w) = \frac{1}{k^2 \rho c^2} \frac{\partial}{\partial z} \tilde{p}(w,z)\big|_{z=0\pm}$$
$$= i \frac{a\tilde{p}_I}{2k^2 \rho c^2} \sum_{m=0}^{\infty} \tau_m \Gamma\left(m + \frac{5}{2}\right) \int_0^{\infty} \left(\frac{2}{\mu a}\right)^{m+(1/2)}$$
$$\times J_{m+(3/2)}(\mu a) J_0(\mu w) \sigma d\mu. \qquad (10)$$

The solution[13] to the infinite integral is given by

$$\tilde{\eta}(w) = i \frac{\tilde{p}_I}{2k^2 a \rho c^2} \sum_{m=0}^{M} \tau_m I_m(w,k), \qquad (11)$$

where

$$I_m(w,k) = I_{mR}(w,k) + iI_{mI}(w,k), \qquad (12)$$

the real part of which is defined by

$$I_{mR}(w,k) = \sqrt{\pi} \sum_{q=0}^{Q} \sum_{r=0}^{R} \frac{(-1)^{q+r} \Gamma(m+5/2)}{(q!)^2 r! \Gamma(r+m+5/2)}$$
$$\times \frac{\Gamma(q+r+1)}{\Gamma(q+r+5/2)} \left(\frac{ka}{2}\right)^{2(q+r)+3} \left(\frac{w}{a}\right)^{2q}, \qquad (13)$$

and the imaginary part is defined by

$$I_{mI}(w,k) = \sqrt{\pi} \sum_{q=0}^{Q} \sum_{r=0}^{R} \frac{(-1)^{q+r+m} \Gamma(m+5/2)}{(q!)^2 r! \Gamma(r-m-1/2)}$$
$$\times \frac{\Gamma(q+r-m-1/2)}{\Gamma(q+r-m+1)} \left(\frac{ka}{2}\right)^{2(q+r-m)} \left(\frac{w}{a}\right)^{2q}. \qquad (14)$$

Let

$$\left(\frac{w}{a}\right)^{2q} = \sum_{n=1}^{\infty} a_n J_0(\alpha_n w/a), \qquad (15)$$

where α_n is the nth zero of $J_0(\alpha_n)$ [i.e., nth solution of $J_0(\alpha_n)=0$]. Multiplying through by the normalizing function $J_0(\alpha_n w/a)$ and integrating over w while applying the property of orthogonality leads to

$$a_n = \frac{\displaystyle\int_0^a (w/a)^{2q} J_0(\alpha_k w/a) w dw}{\displaystyle\int_0^a J_0(\alpha_n w/a) J_0(\alpha_k w/a) w dw}$$
$$= \frac{{}_1F_2(q+1);1, q+2;-\alpha_n^2/4)}{(q+1) J_1^2(\alpha_n)}, \qquad (16)$$

where the following identities[14,15] have been used:

$$\int_0^a J_0(\alpha_n w/a) J_0(\alpha_k w/a) w dw = \begin{cases} a^2 J_1^2(\alpha_n)/2, & \alpha_k = \alpha_n \\ 0, & \alpha_k \ne \alpha_n, \end{cases}, \qquad (17)$$

$$\int_0^a \left(\frac{w}{a}\right)^{2q} J_0(\alpha_k w/a) w dw = \frac{a^2 {}_1F_2(q+1;1,q+2;-\alpha_k^2/4)}{2(q+1)}. \qquad (18)$$

Hence

$$\tilde{\eta}(w) = i \frac{\tilde{p}_I}{2k^2 a \rho c^2} \sum_{m=0}^{M} \tau_m \sum_{q=0}^{Q} ({}_m B_q(ka) + i {}_m S_q(ka))$$
$$\times \sum_{n=1}^{N} \frac{{}_1F_2(q+1;1,q+2;-\alpha_n^2/4)}{(q+1) J_1^2(\alpha_n)} J_0(\alpha_n w/a), \quad (19)$$

where S and B are dipole cylindrical wave functions. S has been named the Streng function[13] as defined by

$${}_m S_q(kb) = \sqrt{\pi} \sum_{r=0}^{R} \frac{(-1)^{q+r+m}(ka/2)^{2(q+r-m)}}{r!(q!)^2(m+5/2)_{r-2m-3}(q+r-m-1/2)_{3/2}} \qquad (20)$$

and B has been named the Bouwkamp function[13] as defined by

$${}_m B_q(kb) = \sqrt{\pi} \sum_{r=0}^{R} \frac{(-1)^{q+r}(ka/2)^{2(q+r)+3}}{r!(q!)^2(m+5/2)_r(q+r+1)_{3/2}}. \qquad (21)$$

where $(x)_n$ is the Pochhammer symbol.

2462   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

T. Mellow and L. Kärkkäinen: Membrane in free space and baffle

2. High frequency approximation

At high frequencies, the following approximation can be used:

$$I_m(w,k) \approx a^2 \Gamma\left(m + \frac{5}{2}\right) \int_0^\infty \left(\frac{2}{a\mu}\right)^{m+(1/2)} J_{m+(3/2)}$$

$$\times (a\mu) J_0(w\mu) k d\mu, \tag{22}$$

the solution[16] to which is given by

$$I_m(w,k) \approx ka\left(m + \frac{3}{2}\right)\left(1 - \frac{w^2}{a^2}\right)^{m+(1/2)}. \tag{23}$$

Let

$$\left(1 - \frac{w^2}{a^2}\right)^{m+(1/2)} = \sum_{n=1}^\infty a_n J_0(\alpha_n w/a), \tag{24}$$

so that multiplying through by the normalizing function $J_0(\alpha_k w/a)$ and integrating over w while applying the property of orthogonality leads to

$$a_n = \frac{\displaystyle\int_0^a (1-(w/a)^2)^{m+(1/2)} J_0(\alpha_k w/a) w dw}{\displaystyle\int_0^a J_0(\alpha_n w/a) J_0(\alpha_k w/a) w dw}$$

$$= \Gamma\left(m + \frac{3}{2}\right)\left(\frac{2}{\alpha_n}\right)^{m+(3/2)} \frac{J_{m+(3/2)}(\alpha_n)}{J_1^2(\alpha_n)}, \tag{25}$$

where Sonine's integral[14] has been used as follows:

$$\int_0^a \left(1 - \frac{w_0^2}{a^2}\right)^{m+(1/2)} J_0(\alpha_n w_0/a) w_0 dw_0$$

$$= \frac{a^2}{2} \Gamma\left(m + \frac{3}{2}\right)\left(\frac{2}{\alpha_n}\right)^{m+(3/2)} J_{m+(3/2)}(\alpha_n) \tag{26}$$

together with Eq. (17). Hence

$$\tilde{\eta}(w) \approx \frac{i\tilde{p}_I}{2k\rho c^2} \sum_{m=0}^\infty \tau_m \Gamma\left(m + \frac{5}{2}\right) \sum_{n=1}^\infty \left(\frac{2}{\alpha_n}\right)^{m+(3/2)}$$

$$\times \frac{J_{m+(3/2)}(\alpha_n)}{J_1^2(\alpha_n)} J_0(\alpha_n w/a). \tag{27}$$

C. Membrane wave equation

The loudspeaker configuration is shown in Fig. 2, with stationary electrodes on either side of the membrane, each at a distance d from it. The steady state wave equation for the membrane deflection $\tilde{\eta}(w)$ can be written with the inherent membrane forces on the left and the external forces on the right:

$$(T\nabla^2 - i\omega z_s - \omega^2 \rho_D h)\tilde{\eta}(w) = \tilde{p}_+(w) - \tilde{p}_-(w) - \tilde{p}_I, \tag{28}$$

where
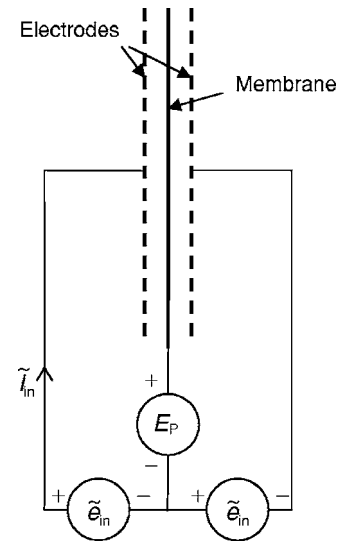
FIG. 2. Electrostatic loudspeaker configuration.

$$\nabla^2 = \frac{\partial^2}{\partial w^2} + \frac{1}{w}\frac{\partial}{\partial w}, \tag{29}$$

where $\tilde{p}_+(w)$ and $\tilde{p}_-(w)$ are the front and rear pressure distributions, respectively, due to the surrounding acoustic medium, T is the tension (which is evenly distributed throughout the membrane), ρ_D is the density of the membrane material, h is its thickness, S is its surface area (given by $S = \pi a^2$), and z_s is an arbitrary specific acoustic impedance which is defined by

$$z_S = R_S + ikcM_S - \frac{C_{ED}}{ikcS}\left(\frac{E_P}{d}\right)^2, \tag{30}$$

where R_S and M_S are the distributed resistance and mass, respectively, of the perforations in the electrodes on either side of the membrane, which are usually designed to damp its vibration modes[17,18]. For the purpose of this analysis, it is assumed that the mass is negligible and that the resistance is linear and will not vary with frequency. The remaining term is the "negative impedance"[1] that results from the increase in electrostatic attraction toward each electrode as the membrane approaches it. The capacitance C_{ED} between the electrodes is given by

$$C_{ED} = \varepsilon_0 S/d, \tag{31}$$

where ε_0 is the permittivity of free space. Due to the push-pull arrangement, it is assumed that the change in capacitance on each side of the membrane, as it is displaced, will tend to balance out. Hence the electrostatic forces will hardly vary, providing there is enough tension. The total driving force $\tilde{F}_I$ is related to the driving pressure $\tilde{p}_I$ and input voltage $\tilde{e}_{in}$ by

$$\tilde{F}_I = S\tilde{p}_I = 2C_{ED}\frac{E_P}{d}\tilde{e}_{in}, \tag{32}$$

where E_P is the membrane polarizing voltage and d is the membrane to electrode spacing. Whilst applying Eq. (1) to the rear pressure term, Eq. (28) can be written in the "Helm-

holtz" form using the modified diaphragm wave number k_D' as follows:

$$(\nabla^2 + k_D'^2)\tilde{\eta} = \frac{1}{T}(2\tilde{p}_+(w) - \tilde{p}_I), \tag{33}$$

where

$$k_D'(k) = k_D(k)\sqrt{1 - i\frac{z_S}{k_D(k)\sqrt{\rho_D h T}}}, \tag{34}$$

$$k_D(k) = \omega/c_D = kc/c_D, \tag{35}$$

and

$$c_D = \sqrt{\frac{T}{\rho_D h}}, \tag{36}$$

where c_D is the speed of sound in the membrane.

D. Solution of the dynamic membrane wave equation

The solution to membrane wave equation, Eq. (33), subject to the edge constraint of Eq. (5), is essentially a two-dimensional counterpart to the Kirchhoff-Helmholtz volume integral as follows:

$$\tilde{\eta}(w) = \frac{1}{T}\int_0^{2\pi}\int_0^a (2\tilde{p}_+(w_0) - \tilde{p}_I)G(w|w_0)w_0 dw_0 d\phi_0. \tag{37}$$

The Green's function[7] for the membrane can be written using the modified wave number k_D' and suppressing the axial term in ϕ and ϕ_0 as follows:

$$G(w|w_0) = \frac{1}{\pi}\sum_{n=1}^{\infty} \frac{J_0(\alpha_n w/a)J_0(\alpha_n w_0/a)}{J_1^2(\alpha_n)(\alpha_n^2 - k_D'^2 a^2)}, \quad 0 \leq w \leq a, \tag{38}$$

where α_n is the nth zero of $J_0(k_D'a)$ such that $J_0(k_D'a)=0$ when $k_D'a=\alpha_n$. Inserting Eqs. (6), (9), and (38) in Eq. (37) and integrating over the surface of the membrane and baffle yields

$$\tilde{\eta}(w) = -\frac{a^2}{T}\tilde{p}_I\sum_{n=1}^{\infty} \frac{J_0(\alpha_n w/a)}{J_1(\alpha_n)(\alpha_n^2 - k_D'^2 a^2)}\left(\frac{2}{\alpha_n} - \sum_{m=0}^{\infty}\tau_m\Gamma\left(m + \frac{5}{2}\right)\right.$$
$$\left.\times\left(\frac{2}{\alpha_n}\right)^{m+(3/2)}\frac{J_{m+(3/2)}(\alpha_n)}{J_1(\alpha_n)}\right), \tag{39}$$

where the following identity[14] has been used:

$$\int_0^a J_0(\alpha_k w_0/a)w_0 dw_0 = a^2 J_1(\alpha_k)/\alpha_k, \tag{40}$$

together with Eq. (26).

E. Final set of simultaneous equations for the power series coefficients

1. Rigorous solution

Equating the right-hand sides of Eqs. (19) and (39) and then equating the coefficients of $J_0(\alpha_n w/a)$, yields the following set of $M+1$ simultaneous equations in τ_m:

$$\sum_{m=0}^{M} {}_m\Psi_n(k_D'a,ka)\tau_m = 1, \quad n = 1,2,\ldots,M+1, \tag{41}$$

where

$${}_m\Psi_n(k_D'a,ka) = \frac{\alpha_n}{2J_1(\alpha_n)}\left\{\Gamma\left(m + \frac{5}{2}\right)\left(\frac{2}{\alpha_n}\right)^{m+(3/2)}\right.$$
$$\times J_{m+(3/2)}(\alpha_n) - \frac{\alpha_n^2 - k_D'^2 a^2}{\alpha^2(ka)}$$
$$\times\sum_{q=0}^{Q}({}_m\mathbf{S}_q(ka) + i{}_m\mathbf{B}_q(ka))$$
$$\left.\times\frac{{}_1F_2(q+1;1,q+2;-\alpha_n^2/4)}{q+1}\right\} \tag{42}$$

and

$$\alpha(ka) = a\omega\sqrt{\frac{2a\rho}{T}} = ka\sqrt{\frac{2a\gamma P_0}{T}}, \tag{43}$$

where the infinite series have been truncated to orders M, Q, and R. $\mathbf{S}$ and $\mathbf{B}$ are dipole cylindrical wave functions defined in Eqs. (20) and (21), respectively. The dimensionless parameter α is the fluid-loading factor, where P_0 is the static pressure defined by $P_0 = \rho c^2/\gamma$ and γ is the adiabatic index. The calculations were performed using, in the case of the damped membrane, 50 digit precision with $M=10+ka$ and $Q=R=2M$. In the case of the undamped membrane, 100 digit precision was used with $M=10+5ka$ and again $Q=R=2M$.

2. High frequency approximation

For large values of ka, the following high-frequency approximation can be obtained by equating the right-hand sides of Eqs. (27) and (39) and then equating the coefficients of $J_0(\alpha_{0n}w/a)$ so that

$${}_m\Psi_n(k_D'a,ka) \approx \Gamma\left(m + \frac{5}{2}\right)\left(\frac{2}{\alpha_n}\right)^{m+(1/2)}\frac{J_{(m+3/2)}(\alpha_n)}{J_1(\alpha_n)}$$
$$\times\left\{1 - \frac{ika(\alpha_n^2 - k_D'^2 a^2)}{\alpha^2(ka)}\right\}. \tag{44}$$

F. Impedance and efficiency

1. Input impedance

The total volume velocity $\tilde{U}_0$ produced by the membrane is equal to the integral of its velocity $\tilde{u}_0(w)(=i\omega\tilde{\eta}(w))$ from Eq. (10) over its surface as follows:

$$\widetilde{U}_0 = ikc \int_0^{2\pi} \int_0^a \widetilde{\eta}(w)w\,dw\,d\phi$$

$$= \frac{a\widetilde{F}_I}{2k\rho c} \sum_{m=0}^{M} \tau_m \Gamma\left(m + \frac{5}{2}\right) \int_0^\infty \left(\frac{2}{a\mu}\right)^{m+3/2}$$

$$\times J_{m+3/2}(a\mu)J_1(a\mu)\sqrt{k^2 - \mu^2}\,d\mu, \tag{45}$$

where the identity of Eq. (40) has again been used. Finally, after solving the infinite integral[15]

$$\widetilde{U}_0 = \frac{\widetilde{F}_I}{2\rho c} \sum_{m=0}^{M} \tau_m \left\{ 1 - {}_2F_3\left(\frac{m}{2} + \frac{3}{4}, \frac{m}{2}\right.\right.$$
$$\left. + \frac{5}{4}; \frac{3}{2}, m + \frac{3}{2}, m + \frac{5}{2}; -k^2a^2\right)$$
$$\left. - i\frac{{}_2F_3\left(\frac{m}{2} + \frac{1}{4}, \frac{m}{2} + \frac{3}{4}; \frac{1}{2}, m+1, m+2; -k^2a^2\right)}{ka(m + 1/2)_{1/2}(m + 5/2)_{-1/2}} \right\}. \tag{46}$$

The mechanical motional impedance z_{mi} is simply the ratio of the applied force $\widetilde{F}_I$ to the average membrane velocity. It can also be expressed in terms of the volume velocity

$$z_{mi} = \frac{\widetilde{F}_I}{\widetilde{u}_0} = \frac{S\widetilde{F}_I}{\widetilde{U}_0} = 2S\rho cZ_I, \tag{47}$$

where the normalized motional input impedance Z_I is given by

$$Z_I = \frac{\widetilde{F}_I}{2\rho c\widetilde{U}_0}. \tag{48}$$

Then z_E is the electrical input impedance, which is given by

$$z_E = \frac{z_{ES}z_{EM}}{z_{ES} + z_{EM}}, \tag{49}$$

where z_{ES} is the *static* electrical impedance given by

$$z_{ES} = \frac{1}{i\omega C_{ED}} = \frac{d}{i\omega\varepsilon_0 S} \tag{50}$$

and z_{EM} is the *motional* electrical impedance given by

$$z_{EM} = \frac{z_{mi}}{2(C_{ED}E_P/d)^2}. \tag{51}$$

The input impedance for a damped membrane is shown in Fig. 3. Since the static impedance dominates, the motional impedance is plotted separately in Fig. 4.

2. Radiation impedance

The total radiation force $\widetilde{F}_R$ acting upon the membrane can be found by integrating the surface pressure from Eq. (6) over its surface as follows:

FIG. 3. Electrical input impedance of push-pull electrostatic loudspeaker in free space where $a=250$ mm, $h=12$ μm, $T=700$ N/m, $\rho_D=1400$ kg/m^3, $E_P=3000$ V, $d=2$ mm, $r=1$ m and $R_S=100$ Ns/m^3.

$$\widetilde{F}_R = -\int_0^{2\pi} \int_0^a \widetilde{p}_+(w_0)w_0\,dw_0\,d\phi_0 = -\frac{\widetilde{F}_I}{2} \sum_{m=0}^{M} \tau_m. \tag{52}$$

Let a membrane force transmission coefficient ζ be defined by

$$\zeta = -2\widetilde{F}_R/\widetilde{F}_I \tag{53}$$

or simply the ratio of the total radiation impedance (on both sides) to the motional impedance. Then from Eq. (52)

$$\zeta = \sum_{m=0}^{M} \tau_m. \tag{54}$$

The acoustic radiation impedance z_{ar} is then given by

$$z_{ar} = \frac{\widetilde{F}_R}{S\widetilde{U}_0} = \frac{\rho c}{S}(R_R + iX_R), \tag{55}$$

where R_R is the normalized radiation *resistance* given by $R_R = \Re(\zeta Z_I)$ and X_R is the normalized radiation *reactance* given by $X_R = \Im(\zeta Z_I)$. This result is plotted in Fig. 5 together with the radiation impedance of a rigid disk in free space for comparison.
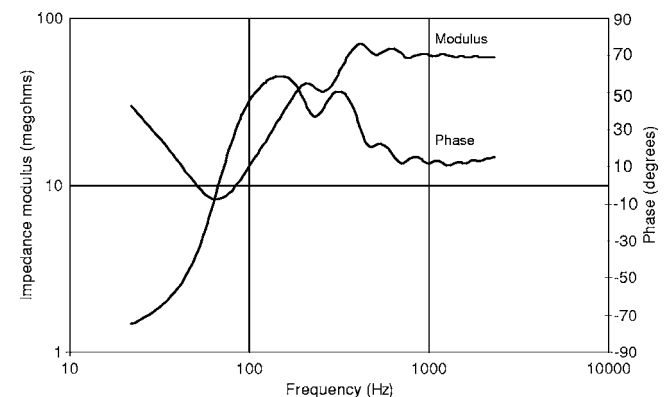
FIG. 4. Motional part of electrical impedance in free space with the same parameters as in Fig. 3.
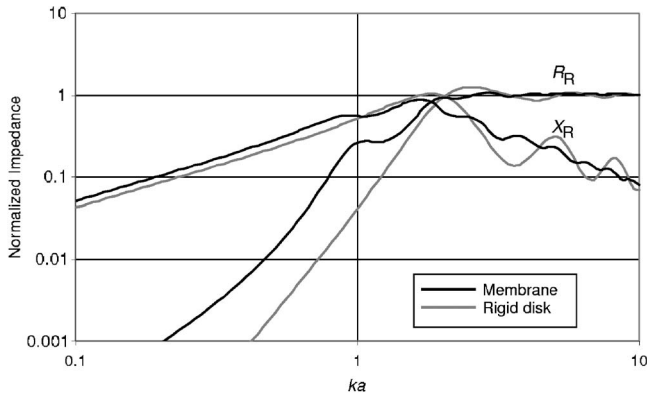
FIG. 5. Normalized radiation impedance in free space with same parameters as in Fig. 3 and radiation impedance of a rigid disk in free space.

3. Efficiency

Now it is easy to calculate the efficiency which is simply equal to $100^* R_R/R_I$ and is shown in Fig. 10 along with the far-field on-axis pressure response and peak displacement.

4. Diaphragm impedance

The mechanical diaphragm impedance z_{md} is the difference between the input impedance and the radiation impedance on both sides where

$$z_{md} = \frac{\widetilde{F}_I - 2\widetilde{F}_R}{\widetilde{u}_0} = \frac{S(\widetilde{F}_I - 2\widetilde{F}_R)}{\widetilde{U}_0} = 2S\rho c(R_D + iX_D), \quad (56)$$

where R_D is the normalized diaphragm *resistance* given by $R_D = \Re(Z_I(1-\zeta))$ and X_D is the normalized membrane *reactance* given by $X_D = \Im(Z_I(1-\zeta))$. This result is plotted in Fig. 6. The various impedances are shown on an equivalent electrical circuit in Fig. 7.

G. Near-field pressure response

1. Near-field pressure as an integral expression

After truncating the summation limit, Eq. (7) can be separated into finite and infinite integrals as follows:
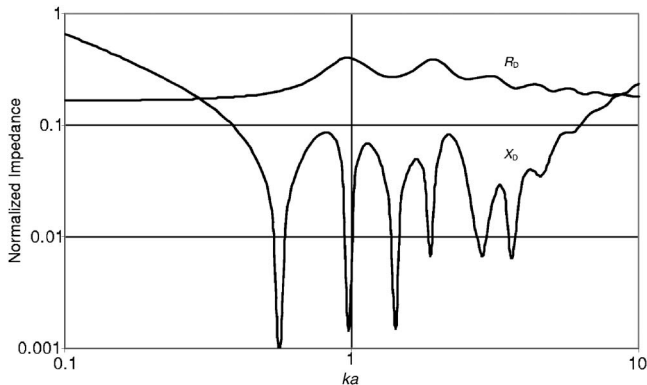
FIG. 6. Normalized diaphragm impedance in free space with the same parameters as in Fig. 3.
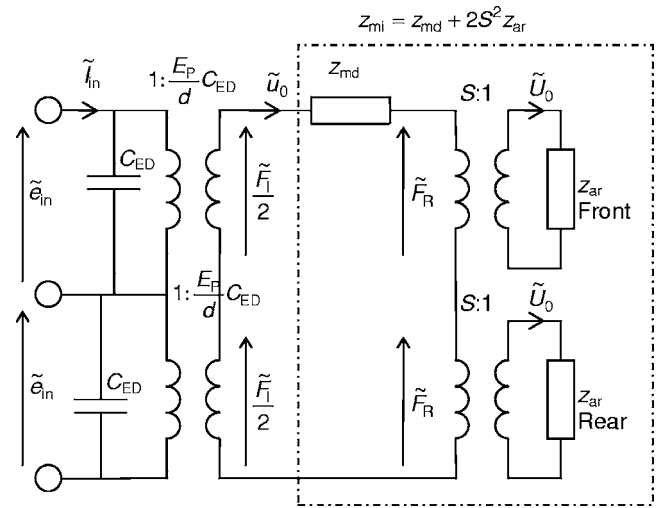
FIG. 7. Equivalent electrical circuit.

$$\widetilde{p}(w,z) = -\frac{\widetilde{p}_I}{2}\sum_{m=0}^{M}\tau_m\Gamma\left(m + \frac{5}{2}\right)(I_{\text{Fin}}(m,w,z)$$

$$+ I_{\text{Inf}}(m,w,z)), \quad (57)$$

where

$$I_{\text{Fin}}(m,w,z) = a\int_0^k \left(\frac{2}{a\mu}\right)^{m+(1/2)} J_{m+(3/2)}(a\mu)$$

$$\times J_0(w\mu)e^{-iz\sqrt{k^2-\mu^2}}d\mu \quad (58)$$

and

$$I_{\text{Inf}}(m,w,z) = a\int_k^\infty \left(\frac{2}{a\mu}\right)^{m+(1/2)} J_{m+(3/2)}(a\mu)$$

$$\times J_0(w\mu)e^{-z\sqrt{\mu^2-k^2}}d\mu, \quad (59)$$

2. Solution of the finite integral

Substituting $\mu = k\sqrt{1-t^2}$ in Eq. (58) in order to simplify the exponent yields

$$I_{\text{Fin}}(m,w,z) = 2\left(\frac{2}{ka}\right)^{m-(1/2)}$$

$$\times \int_0^1 \frac{J_{m+(3/2)}(ka\sqrt{1-t^2})J_0(kw\sqrt{1-t^2})}{(1-t^2)^{(m/2)+(3/4)}}e^{-ikzt}tdt. \quad (60)$$

The Bessel functions in Eq. (60) can then be expanded using the following Lommel expansion:[19]

$$\frac{J_n(ka\sqrt{1-t^2})}{(1-t^2)^{n/2}} = \sum_{m=0}^\infty \left(\frac{ka}{2}\right)^m \frac{t^{2m}}{m!}J_{n+m}(ka), \quad (61)$$

which leads to

T. Mellow and L. Kärkkäinen: Membrane in free space and baffle

$$I_{\text{Fin}}(m,w,z) = 2\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\left(\frac{ka}{2}\right)^{p-m+(1/2)}\left(\frac{kw}{2}\right)^{q}$$

$$\times\frac{J_{p+m+(3/2)}(ka)J_{q}(kw)}{p!q!}\int_{0}^{1}e^{-ikzt}t^{2(p+q)+1}dt. \tag{62}$$

The integral in Eq. (62) can be solved using the identity[14]

$$\int_{0}^{1}e^{-ikzt}t^{2(p+q)+1}dt = \frac{\gamma(2p+2q+2,ikz)}{(ikz)^{2(p+q)+2}}, \tag{63}$$

where γ is the incomplete gamma function. Inserting Eq. (63) in Eq. (62) and truncating the summation limits gives the final solution to Eq. (60) as follows:

$$I_{\text{Fin}}(m,w,z) = 2\sum_{p=0}^{P}\sum_{q=0}^{Q}\frac{1}{p!q!(ikz)^{2(p+q+1)}}\left(\frac{ka}{2}\right)^{p-m+(1/2)}$$

$$\times\left(\frac{kw}{2}\right)^{q}J_{p+m+(3/2)}(ka)J_{q}(kw)$$

$$\times\gamma(2p+2q+2,ikz). \tag{64}$$

3. Solution of the infinite integral

a. Expansion of the Bessel functions. Substituting $\mu = k\sqrt{t^2+1}$ in Eq. (59) in order to simplify the exponent yields

$$\frac{J_n(ka\sqrt{t^2+1})}{(t^2+1)^{n/2}} = \begin{cases} 2\sum_{m=0}^{\infty}\dfrac{(-1)^m}{1+\delta_{0m}}J_{2m}(ka)J_{2m}(kat), & n=0 \\[2ex] \left(\dfrac{2}{kat}\right)^n\sum_{m=0}^{\infty}\dfrac{(-1)^m}{m!}(2m+n)\Gamma(m+n)\times J_{2m+n}(ka)J_{2m+n}(kat), & n\neq 0, \end{cases} \tag{69}$$

where δ_{0m} is the Kronecker delta function. If both Bessel functions were to be expanded (in a procedure similar to that of Williams[20]) using the Lommel expansion of Eq. (66), together with the following identities[14]

$$\int_{0}^{\infty}e^{-kzt}t^{2(p+q)+1}dt = \frac{\Gamma(2p+2q+2)}{(kz)^{2(p+q)+2}}, \tag{70}$$

$$\gamma(2p+2q+2,ikz) - \Gamma(2p+2q+2)$$
$$= -\Gamma(2p+2q+2,ikz), \tag{71}$$

the resulting solution

$$I_{\text{Fin}} + I_{\text{Inf}} = 2\sum_{p=0}^{P}\sum_{q=0}^{Q}\frac{1}{p!q!(ikz)^{2(p+q+1)}}\left(\frac{ka}{2}\right)^{p-m+(1/2)}\left(\frac{kw}{2}\right)^{q}$$

$$\times J_{p+m+(3/2)}(ka)J_{q}(kw)\Gamma(2p+2q+2,ikz) \tag{72}$$

$$I_{\text{Inf}}(m,w,z) = 2\left(\frac{2}{ka}\right)^{m-(1/2)}\int_{0}^{\infty}\frac{J_{m+(3/2)}(ka\sqrt{t^2+1})}{(t^2+1)^{(m/2)+(3/4)}}$$

$$\times J_0(kw\sqrt{t^2+1})e^{-kzt}tdt. \tag{65}$$

The Bessel functions in Eq. (65) can expanded in two possible ways. The first is the Lommel expansion[19] as follows:

$$\frac{J_n(ka\sqrt{t^2+1})}{(t^2+1)^{n/2}} = \sum_{m=0}^{\infty}(-1)^m\left(\frac{ka}{2}\right)^m\frac{t^{2m}}{m!}J_{m+n}(ka), \tag{66}$$

and the second is Gegenbauer's summation theorem[14]

$$\frac{J_n(ka\sqrt{t^2+1})}{(t^2+1)^{n/2}} = \Gamma(n)\left(\frac{2}{kat}\right)^n\sum_{m=0}^{\infty}(m+n)J_{m+n}(kat)$$

$$\times J_{m+n}(ka)C_m^n(0), \tag{67}$$

where C_m^n is the Gegenbauer[14] polynomial given by

$$C_m^n(0) = \begin{cases} 1, & m=n=0 \\[2ex] \dfrac{\cos(m\pi/2)\Gamma(n+m/2)}{(m/2)!\Gamma(n)}, & m\neq 0 \text{ or } n\neq 0, \end{cases} \tag{68}$$

where n is a positive real integer. Inserting Eq. (68) in Eq. (67) and noting that, due to the cosine term, all odd terms of the Gegenbauer polynomial are zero, yields

would only converge for $z^2 \geq a^2 + w^2$ and would therefore be of limited value. On the other hand, expanding both Bessel functions using the summation theorem leads to an integral of the form

$$\int_{0}^{\infty}J_{2p+m+(3/2)}(kat)J_{2q}(kwt)e^{-kzt}t^{-1}dt, \tag{73}$$

which can be solved, but the solution is in the form of an expansion and is therefore slow to compute. Hence, the following solution uses the Lommel expansion for one Bessel function and the summation theorem for the other and then exchanges them so that convergent solutions are obtained, in turn, for $0 < w^2 + z^2 \leq a^2$ and $w^2 + z^2 \geq a^2$ (providing $z > 0$).

b. Solution when the distance from the center of the membrane to the observation point is greater than the membrane's radius Expanding $J_{m+(3/2)}(ka\sqrt{t^2+1})$ in Eq. (65) with Eq. (66) and $J_0(kw\sqrt{t^2+1}))$ with Eq. (69) yields

$$I_{\text{Inf}}(m,w,z) = 4\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{(-1)^{p+q}}{p!(1+\delta_{0q})}\left(\frac{ka}{2}\right)^{p-m+(1/2)}$$

$$\times J_{p+m+(3/2)}$$

$$\times (ka)J_{2q}(kw)\int_0^{\infty}e^{-kzt}J_{2q}(kwt)t^{2p+1}dt. \quad (74)$$

The integral in Eq. (74) can be solved using the following identity:[14]

$$\int_0^{\infty}e^{-kzt}J_{2q}(kwt)t^{2p+1}dt = \frac{\Gamma(2p+2q+2)}{(k^2z^2+k^2w^2)^{p+1}}$$

$$\times P_{2p+1}^{-2q}\left(\frac{z}{\sqrt{z^2+w^2}}\right), \quad (75)$$

so that, after truncating the summation limits, the final solution is given by

$$I_{\text{Inf}}(m,w,z) = 4\sum_{p=0}^{P}\sum_{q=0}^{Q}\frac{(-1)^{p+q}\Gamma(2p+2q+2)}{(1+\delta_{0q})p!(k^2z^2+k^2w^2)^{p+1}}$$

$$\times\left(\frac{ka}{2}\right)^{p-m+(1/2)}J_{p+m+(3/2)}(ka)J_{2q}(kw)$$

$$\times P_{2p+1}^{-2q}\left(\frac{z}{\sqrt{z^2+w^2}}\right). \quad (76)$$

This expression converges for $w^2+z^2\geq a^2$ and can therefore be used in that region, providing $z>0$.

c. Solution when the distance from the center of the membrane to the observation point is less than the membrane's radius Expanding $J_{m+(3/2)}(ka\sqrt{t^2+1})$ in Eq. (65) with Eq. (69) and $J_0(kw\sqrt{t^2+1}))$ with Eq. (66) yields

$$I_{\text{Inf}}(m,w,z) = 2\left(\frac{2}{ka}\right)^{2m+1}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{(-1)^{p+q}\Gamma(p+m+3/2)}{p!q!}$$

$$\times(2p+m+3/2)\left(\frac{kw}{2}\right)^q J_{2p+m+(3/2)}(ka)$$

$$\times J_q(kw)\int_0^{\infty}e^{-kzt}J_{2p+m+(3/2)}(kat)t^{2q-m-(1/2)}dt. \quad (77)$$

The integral in Eq. (77) can be solved using the identity of Eq. (75) so that, after truncating the summation limits, the final solution is given by

$$I_{\text{Inf}}(m,w,z) = 2\left(\frac{2}{ka}\right)^{2m+1}\sum_{p=0}^{P}\sum_{q=0}^{Q}\frac{(-1)^{p+q}(2p+m+3/2)}{q!(k^2z^2+k^2a^2)^{q-(m/2)+(1/4)}}$$

$$\times(p+1)_{m+(1/2)}\Gamma(2p+2q+2)\left(\frac{kw}{2}\right)^q$$

$$\times J_{2p+m+(3/2)}(ka)J_q(kw)P_{2q-m-(1/2)}^{-2p-m-(3/2)}\left(\frac{z}{\sqrt{z^2+a^2}}\right). \quad (78)$$

This expression converges for $w^2<a^2+z^2$ and can therefore be used in the region $0<w^2+z^2<a^2$, providing $z>0$. The normalized near-field pressure response of the membrane is shown in Fig. 8 for various values of ka. The calculations

FIG. 8. Near-field pressure response in free space with the same parameters as in Fig. 3.

were performed using 50 digit precision with $M=P=50$ and $Q=100$.

H. Far-field pressure response

The far-field pressure is derived using the same procedure as shown in Part K of Sec. II of a recent paper [13] to give
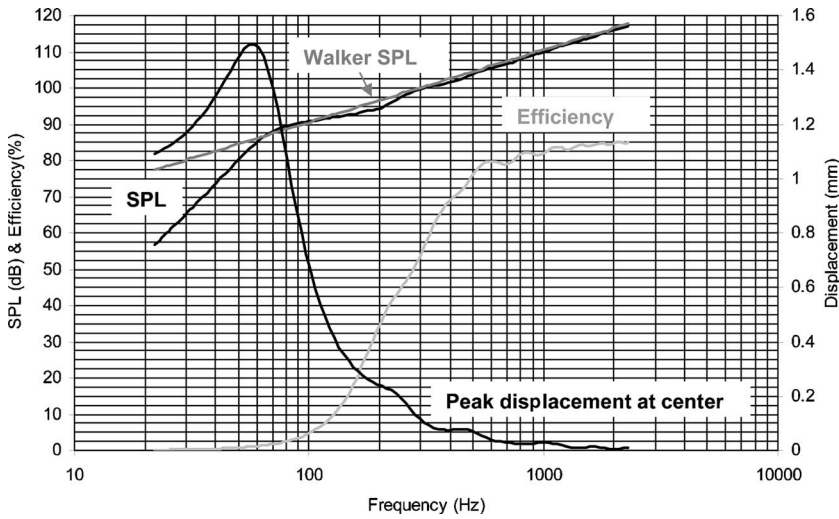
FIG. 9. On-axis pressure response of push-pull electrostatic loudspeaker in free space where $a=250$ mm, $h=12$ μm, $T=700$ N/m, $\rho_D=1400$ kg/m^3, $e_{\mathrm{in}}=2000$ V$_{\mathrm{rms}}$, $E_P=3000$ V, $d=2$ mm, $r=1$ m, and $R_S=100$ Ns/m^3.

$$\widetilde{p}(r,\theta) = -i\frac{a\widetilde{p}_I}{4r}e^{-ikr}D(\theta) = -i\frac{E_P}{d}C_{\mathrm{ED}}\frac{a\widetilde{e}_{\mathrm{in}}}{2rS}e^{-ikr}D(\theta), \quad (79)$$

where the directivity function $D(\theta)$ is given by

$$D(\theta) = ka\cos\theta \sum_{m=0}^{M}\tau_m\Gamma\left(m+\frac{5}{2}\right)\left(\frac{2}{ka\sin\theta}\right)^{m+(3/2)}$$

$$\times J_{m+3/2}(ka\sin\theta), \quad (80)$$

which, for $\theta=0$ (i.e., on-axis), simplifies to

$$D(0) = ka\sum_{m=0}^{M}\tau_m = ka\zeta, \quad (81)$$

using ζ from Eq. (54). (The on-axis response of a resilient disk in free space is given by setting $\zeta=1$.) The on-axis pressure then simplifies to

$$\widetilde{p}(r,0) = -i\frac{ka^2}{4r}\widetilde{p}_I e^{-ikr}\zeta = -i\omega\frac{E_P}{d}C_{\mathrm{ED}}\frac{\widetilde{e}_{\mathrm{in}}}{2\pi rc}e^{-ikr}\zeta. \quad (82)$$

Taking into account that $\widetilde{I}_{\mathrm{in}}\approx i\omega C_{\mathrm{ED}}\widetilde{e}_{\mathrm{in}}$, leads to

$$\widetilde{p}(r,0)|_{b=a} \approx -\zeta\frac{E_P}{d}\cdot\frac{\widetilde{I}_{\mathrm{in}}}{c}\cdot\frac{e^{-ikr}}{2\pi r}, \quad (83)$$

which, in the case of $\zeta=1$, is Walker's equation.[21] A better approximation is given by

$$\zeta \approx 2\rho c/(R_S+2\rho c), \quad (84)$$

which leads to what could be termed the "modified Walker's equation." The on-axis pressure response with damping is shown in Fig. 9, together with the efficiency and peak displacement at $w=0$. Figure 10 shows a comparison of the on-axis responses with and without damping (the level is reduced in order to accommodate the large undamped excursions), together with the high frequency approximation, where the SPL is given by: SPL$=20\log_{10}|\widetilde{p}(r,0)/(20\times10^{-6})|$. The normalized directivity function $20\log_{10}|D(\theta)|/|D(0)|$ is plotted in Fig. 11 for various values of ka.

III. MEMBRANE IN AN INFINITE BAFFLE

A. Boundary conditions

The membrane shown in Fig. 1 is now mounted in an infinite baffle in the xy plane with its center at the origin and the uniform driving force $\widetilde{p}_I$ is applied in the z direction. The membrane deflection $\widetilde{\eta}(w)$ is then to be used as a parameter to couple it to the surrounding loss-free acoustic medium.

The monopole source elements and their images together form the membrane source. Since they are coincident in the plane of the infinite baffle, they combine to form elements of double strength. Hence our membrane may be modeled as a "breathing" membrane in free space. Due to the symmetry of the pressure fields on either side of the plane of symmetry, we have the following Neumann boundary condition on the front and rear surfaces of the infinite baffle:

$$\frac{\partial}{\partial z}\widetilde{p}(w,z)|_{z=0\pm} = 0, \quad a < w \leqslant \infty. \quad (85)$$

Also, on the front and rear outer surfaces of the membrane, there is the coupling condition

$$\frac{\partial}{\partial z}\widetilde{p}(w,z)|_{z=0\pm} = -ik\rho c\widetilde{u}_0(w),$$

$$= k^2\rho c^2\eta(w) \quad 0 \leqslant w \leqslant a. \quad (86)$$

FIG. 10. On-axis pressure response with the same parameters as in Fig. 9, except $e_{\mathrm{in}}=40$ V$_{\mathrm{rms}}$, and with $R_S=0$ (undamped).
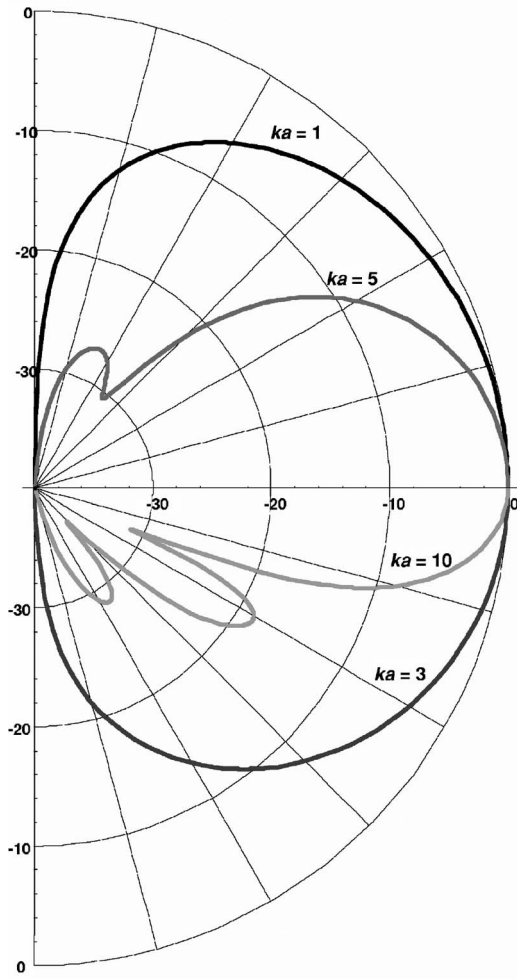
FIG. 11. Normalized far-field directivity function in free space with same parameters as in Fig. 9.

In the actual physical system (as opposed to the "breathing membrane" model), the pressure field on one side of the xy plane is the symmetrical "negative" of that on the other, so that

$$\tilde{p}(w,z) = -\tilde{p}(w,-z). \tag{87}$$

Again, the perimeter of the membrane is fixed, which leads to the boundary condition of Eq. (5). On the surface of the membrane, the velocity distribution $\tilde{u}_0(w_0)$ is defined, according to Bouwkamp's solution[6] to the free space wave equation in oblate spheroidal coordinates, as

$$\tilde{u}_0(w_0) = \sum_{m=0}^{\infty} \tilde{A}_m \left(1 - \frac{w_0^2}{a^2}\right)^{m+(1/2)}, \tag{88}$$

where $\tilde{A}_m$ are the as yet unknown power series coefficients.

B. Solution to the free space wave equation

Using the King integral and taking into account the double layer source, the pressure distribution is defined by

$$\tilde{p}(w,z) = 2\int_0^{2\pi}\int_0^a g(w,z|w_0,z_0)$$

$$\times \frac{\partial}{\partial z_0}\tilde{p}(w_0,z_0)\big|_{z_0=0+}w_0 dw_0 d\phi_0, \tag{89}$$

where the Green's function[14] is defined, in axisymmetric cylindrical coordinates, by

$$g(w,z|w_0,z_0) = \frac{i}{4\pi}\int_0^{\infty} J_0(\mu w)J_0(\mu w_0)\frac{\mu}{\sigma}e^{-i\sigma|z-z_0|}d\mu, \tag{90}$$

where σ is defined by Eq. (8). Substituting Eqs. (86), (88), and (90) in Eq. (89) and integrating over the surface of the membrane yields

$$\tilde{p}(w,z) = ka\rho c\sum_{m=0}^{\infty}\tilde{A}_m\Gamma\left(m+\frac{3}{2}\right)\int_0^{\infty}\left(\frac{2}{a\mu}\right)^{m+(1/2)}$$

$$\times J_0(w\mu)J_{m+(3/2)}(a\mu)\frac{e^{-i\sigma z}}{\sigma}d\mu, \tag{91}$$

where Sonine's integral of Eq. (26) has been used (with $\mu = \alpha_{0n}/a$). Setting $z=0$ yields the surface pressure as follows:

$$\tilde{p}_+(w_0) = ka\rho c\sum_{m=0}^{\infty}\tilde{A}_m\Gamma\left(m+\frac{3}{2}\right)\int_0^{\infty}\left(\frac{2}{a\mu}\right)^{m+(1/2)}$$

$$\times J_0(w_0\mu)J_{m+(3/2)}(a\mu)\frac{1}{\sigma}d\mu. \tag{92}$$

C. Formulation of the coupled problem

Let the power series coefficients $\tilde{A}_m$ be related to normalized dimensionless coefficients τ_m by

$$\tilde{A}_m = \frac{\tau_m(m+3/2)}{2\rho c}\tilde{p}_I. \tag{93}$$

Substituting this together with Eq. (92) in Eq. (37) and integrating over ϕ_0 leads to the following coupled equation:

$$\tilde{\eta}(w) = \frac{2\pi}{T}\tilde{p}_I\int_0^a\left(\sum_{m=0}^{\infty}\tau_m I_m(w_0,k) - 1\right)G(w|w_0)w_0 dw_0. \tag{94}$$

The infinite integral I_m is defined by

$$I_m(w_0,k) = ka\Gamma\left(m+\frac{5}{2}\right)\int_0^{\infty}\left(\frac{2}{a\mu}\right)^{m+(1/2)}$$

$$\times J_0(w_0\mu)J_{m+(3/2)}(a\mu)\frac{1}{\sigma}d\mu$$

$$= I_{mR}(w_0,k) + iI_{mI}(w_0,k), \tag{95}$$

where the solution[10] to the real part is given by

$$I_{mR}(w_0,k) = \sqrt{\pi}\sum_{q=0}^{\infty}\sum_{r=0}^{\infty}\frac{(-1)^{q+r}\Gamma(m+5/2)\Gamma(q+r+1)}{(q!)^2 r!\Gamma(r+m+5/2)\Gamma(q+r+3/2)}$$

$$\times\left(\frac{ka}{2}\right)^{2(q+r+1)}\left(\frac{w_0}{a}\right)^{2q}. \tag{96}$$

and the solution[10] to the imaginary part is given by

$$I_{mI}(w_0,k)$$

$$= \sqrt{\pi}\sum_{q=0}^{\infty}\sum_{r=0}^{\infty}\frac{(-1)^{q+r+m+1}\Gamma(m+5/2)\Gamma(q+r-m-1/2)}{(q!)^2 r!\Gamma(r-m-1/2)\Gamma(q+r-m)}$$

$$\times\left(\frac{ka}{2}\right)^{2(q+r-m)-1}\left(\frac{w_0}{a}\right)^{2q}. \tag{97}$$

For large ka, an approximation[16] is given by

$$I_m(w_0,k) \approx (m+3/2)\left(1-\frac{w_0^2}{a^2}\right)^{m+(1/2)}. \tag{98}$$

D. Bessel series expansion of the membrane deflection

Using the power series of Eq. (88) together with Eq. (93), the deflection can also be expressed as

$$\widetilde{\eta}(w) = \frac{\widetilde{u}_0(w)}{i\omega} = \frac{\widetilde{p}_I}{2ik\rho c^2}\sum_{m=0}^{\infty}\tau_m\left(m+\frac{3}{2}\right)\left(1-\frac{w^2}{a^2}\right)^{m+(1/2)},$$

$$0 \leqslant w \leqslant a. \tag{99}$$

Using the Bessel series expansion from Eqs. (24) and (25), Eq. (99) can be written

$$\widetilde{\eta}(w) = \frac{\widetilde{p}_I}{2ik\rho c^2}\sum_{m=0}^{\infty}\tau_m\Gamma\left(m+\frac{5}{2}\right)\sum_{n=1}^{\infty}\left(\frac{2}{\alpha_n}\right)^{m+(3/2)}$$

$$\times\frac{J_{m+(3/2)}(\alpha_n)}{J_1^2(\alpha_n)}\times J_0(\alpha_n w/a), \quad 0 \leqslant w \leqslant a. \tag{100}$$

E. Final set of simultaneous equations for the power series coefficients

1. Rigorous solution

Inserting Eq. (38) in Eq. (94), equating the right-hand sides of Eqs. (94) and (100), while integrating over the surface of the membrane using the identities of Eqs. (18) and (40) and then equating the coefficients of $J_0(\alpha_n w/a)$ yields the following set of $M+1$ simultaneous equations in τ_m;

$$\sum_{m=0}^{M}{}_m\Psi_n(k_D'a,ka)\tau_m = 1, \quad n = 1,2,\dots,M+1, \tag{101}$$

where

$${}_m\Psi_n(k_D'a,ka) = \frac{\alpha_n}{2J_1(\alpha_n)}\left\{-ika\Gamma\left(m+\frac{5}{2}\right)\left(\frac{2}{\alpha_n}\right)^{m+(3/2)}\right.$$

$$\times J_{m+(3/2)}(\alpha_n)\frac{\alpha_n^2 - k_D'^2 a^2}{\alpha^2(ka)}$$

$$+ \sum_{q=0}^{Q}({}_{m+1}\mathbf{P}_q(ka) + i\,{}_{m+1}\mathbf{T}_q(ka))$$

$$\left.\times\frac{{}_1F_2(q+1;1,q+2;-\alpha_n^2/4)}{q+1}\right\}, \tag{102}$$

where $\mathbf{P}$ and $\mathbf{T}$ are monopole cylindrical wave functions. $\mathbf{P}$ has been named the Spence function[10] as defined by

$${}_{m+1}\mathbf{P}_q(ka) = \sqrt{\pi}\sum_{r=0}^{R}\frac{(-1)^{q+r}(ka/2)^{2(q+r+1)}}{r!(q!)^2(m+5/2)_r(q+r+1)_{1/2}} \tag{103}$$

and $\mathbf{T}$ has been named the Stenzel function[10] as defined by

$${}_{m+1}\mathbf{T}_q(ka) = \sqrt{\pi}\sum_{r=0}^{R}\frac{(-1)^{q+r+m+1}(ka/2)^{2(q+r-m)-1}}{r!(q!)^2(m+5/2)_{r-2m-3}(q+r-m-1/2)_{1/2}}, \tag{104}$$

where the infinite series have been truncated to orders M, Q, and R. $\mathbf{P}$ and $\mathbf{T}$ are monopole cylindrical wave functions. The dimensionless parameter α is the fluid-loading factor, as defined by Eq. (43), and $(x)_n$ is the Pochhammer symbol. The calculations were performed using, in the case of the damped membrane, 50 digit precision with $M=10+ka$ and $Q=R=2M$. In the case of the undamped membrane, 100 digit precision was used with $M=10+5ka$ and again $Q=R=2M$.

2. High frequency approximation

For large values of ka, the following high-frequency approximation can be obtained by the same method as for the rigorous solution, except that Eq. (98) is used instead of Eq. (97) and the identity of Eq. (18) is replaced by Eq. (26) so that the solution is, not surprisingly, the same as that given in Eq. (44).

F. Impedance and efficiency

1. Input impedance

The total volume velocity $\widetilde{U}_0$ produced by the membrane is equal to the integral of its velocity $\widetilde{u}_0(w)$ ($=i\omega\widetilde{\eta}(w)$) from Eq. (99) over its surface as follows:

$$\widetilde{U}_0 = ikc\int_0^{2\pi}\int_0^a\widetilde{\eta}(w)w\,dw\,d\phi = \frac{S\widetilde{p}_I}{2\rho c}\sum_{m=0}^{\infty}\tau_m, \tag{105}$$

where $S = \pi a^2$ is the area of the membrane. The parameters z_{mi}, z_E, z_{ES}, and z_{EM} are given in Eqs. (47) and (49)–(51), respectively. The normalized motional input impedance Z_I is given by
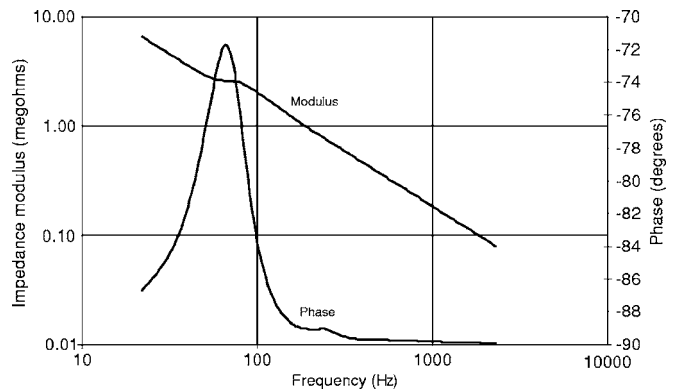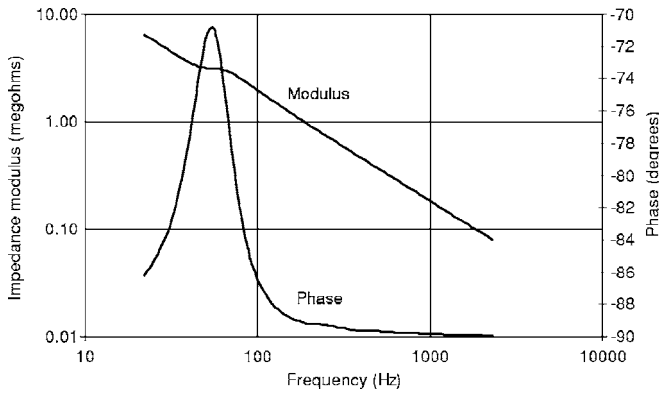
FIG. 12. Electrical input impedance of push-pull electrostatic loudspeaker in infinite baffle where $a=250$ mm, $h=12$ μm, $T=700$ N/m, $\rho_D=1400$ kg/m^3, $E_P=3000$ V, $d=2$ mm, $r=1$ m, and $R_S=100$ Ns/m^3.

$$Z_I = \frac{\widetilde{F}_I}{2\rho c \widetilde{U}_0} = \left(\sum_{m=0}^{M} \tau_m \right)^{-1}. \tag{106}$$

The input impedance for a damped membrane is shown in Fig. 12. Since the static impedance dominates, the motional impedance is plotted separately in Fig. 13.

2. Radiation impedance

The total radiation force $\widetilde{F}_R$ acting upon the membrane can be found by integrating the surface pressure from Eq. (92) over its surface as

$$\widetilde{F}_R = - \int_0^{2\pi} \int_0^a \widetilde{p}_+(w_0) w_0 dw_0 d\phi_0$$

$$= \frac{\widetilde{F}_I}{2} ka \sum_{m=0}^{\infty} \tau_m \Gamma \left(m + \frac{5}{2} \right)$$

$$\times \int_0^{\infty} \left(\frac{2}{a\mu} \right)^{m+(3/2)} \frac{J_1(a\mu) J_{m+(3/2)}(a\mu)}{\sqrt{k^2 - \mu^2}} d\mu, \tag{107}$$
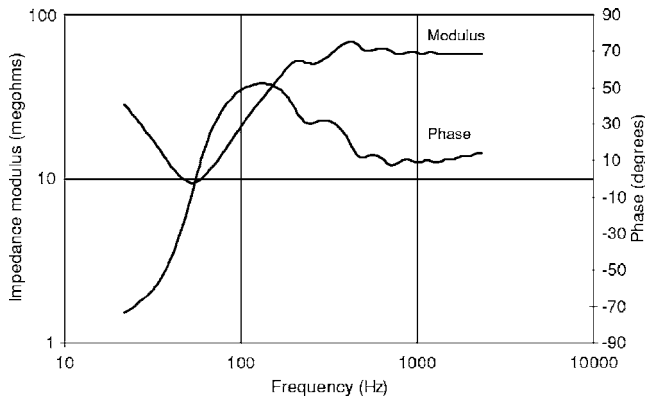
FIG. 13. Motional part of electrical impedance in infinite baffle with the same parameters as in Fig. 12.
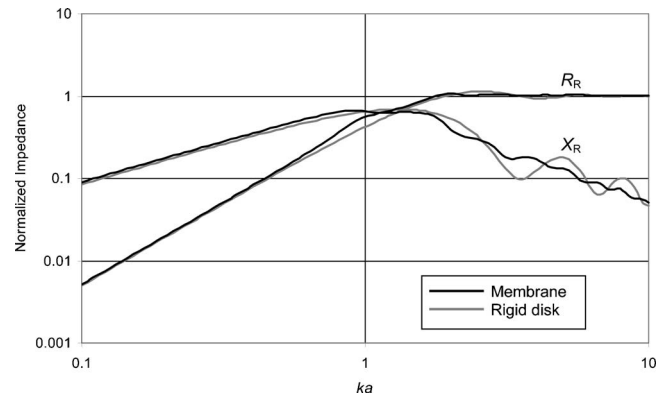
FIG. 14. Normalized radiation impedance in infinite baffle with the same parameters as in Fig. 12 and radiation impedance of a rigid disk in an infinite baffle.

where the identity of Eq. (40) has again been used. After solving the infinite integral,[15] the force transmission coefficient, defined in Eq. (53), is given by

$$\zeta = \sum_{m=0}^{M} \tau_m \left\{ 1 - {}_2F_3 \left(\frac{m}{2} + \frac{3}{4}, \frac{m}{2} + \frac{5}{4}; \frac{1}{2}, m + \frac{3}{2}, m + \frac{5}{2}; -k^2a^2 \right) \right.$$

$$\left. + ika \frac{{}_2F_3 \left(\frac{m}{2} + \frac{5}{4}, \frac{m}{2} + \frac{7}{4}; \frac{3}{2}, m+2, m+3; -k^2a^2 \right)}{(m+5/2)_{-1/2}(m+5/2)_{1/2}} \right\}. \tag{108}$$

The acoustic radiation impedance z_{ar} is then given by Eq. (55), where R_R is the normalized radiation resistance given by $R_R = \Re(\zeta Z_I)$ and X_R is the normalized radiation reactance given by $X_R = \Im(\zeta Z_I)$. This result is plotted in Fig. 14 together with the radiation impedance of a rigid disk in an infinite baffle for comparison.

3. Efficiency

Now it is easy to calculate the efficiency which is simply equal to $100^* R_R/R_I$ and is shown in Fig. 17 along with the far-field on-axis pressure response and peak displacement.

4. Diaphragm impedance

The mechanical diaphragm impedance z_{md} is the difference between the input impedance and the radiation impedance on both sides, as defined by Eq. (56), where R_D is the normalized diaphragm *resistance* given by $R_D = \Re(Z_I(1-\zeta))$ and X_D is the normalized membrane *reactance* given by $X_D = \Im(Z_I(1-\zeta))$. This result is plotted in Fig. 15. The various impedances are shown on an equivalent electrical circuit in Fig. 7.
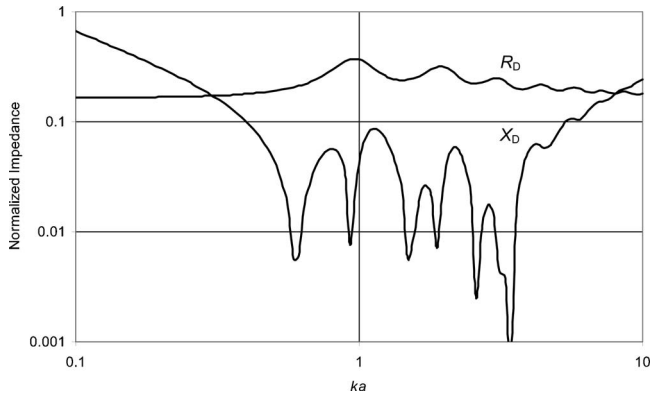
FIG. 15. Normalized diaphragm impedance in infinite baffle with the same parameters as in Fig. 12.

G. Near-field pressure response

1. Near-field pressure when the distance from the center of the membrane to the observation point is greater than the membrane's radius

Using the same procedure as in a recent paper for a resilient disk in an infinite baffle,[10] it can be shown that

$$
p(r,\theta) = -\pi\tilde{p}_I \sum_{m=0}^{M} \sum_{p=0}^{P} \frac{\tau_m h_{2p}^{(2)}(kr) P_{2p}(\cos\theta)}{(m+(5/2))_p \Gamma((1/2)-p)\Gamma(2p+(1/2))}
$$

$$
\times \left(\frac{ka}{2}\right)^{2p+2} {}_1F_2\left(p+1; p+m+\frac{5}{2}, 2p+\frac{3}{2}; -\frac{k^2a^2}{4}\right).
$$

(109)

This solution converges providing $r \geq a$.

2. Near-field pressure when the distance from the center of the membrane to the observation point is less than the membrane's radius

a. Near-field pressure as an integral expression. After truncating the summation limit and inserting Eq. (93), Eq. (91) can be separated into finite and infinite integrals as follows:

$$
\tilde{p}(w,z) = \tilde{p}_I \sum_{m=0}^{M} \tau_m \Gamma\left(m+\frac{5}{2}\right)(I_{\text{Fin}}(m,w,z) - iI_{\text{Inf}}(m,w,z)),
$$

(110)

where

$$
I_{\text{Fin}}(m,w,z) = \frac{ka}{2}\int_0^k \left(\frac{2}{a\mu}\right)^{m+(1/2)} J_{m+(3/2)}(a\mu)
$$

$$
\times J_0(w\mu)\frac{e^{-iz\sqrt{k^2-\mu^2}}}{\sqrt{k^2-\mu^2}}d\mu
$$

(111)

and

$$
I_{\text{Inf}}(m,w,z) = \frac{ka}{2}\int_k^\infty \left(\frac{2}{a\mu}\right)^{m+(1/2)} J_{m+(3/2)}(a\mu)
$$

$$
\times J_0(w\mu)\frac{e^{-z\sqrt{\mu^2-k^2}}}{\sqrt{\mu^2-k^2}}d\mu.
$$

(112)
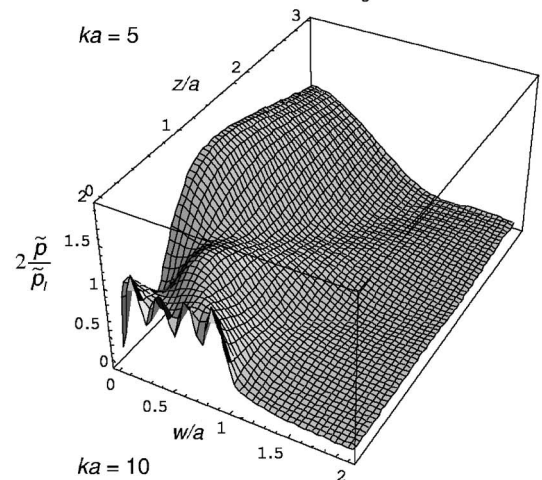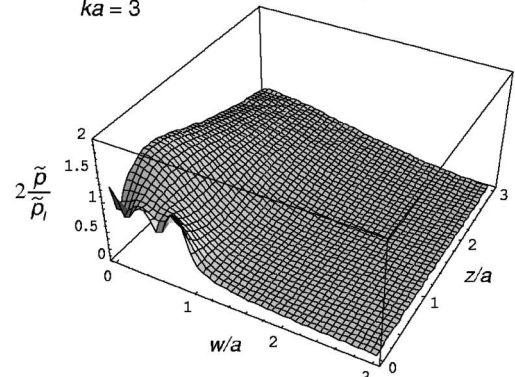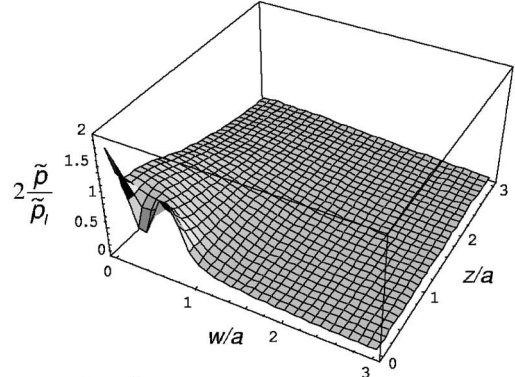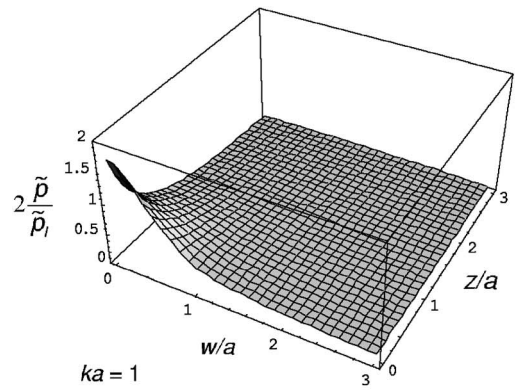
FIG. 16. Near-field pressure response in infinite baffle with the same parameters as in Fig. 12.

b. Solution of the finite integral. Using the same procedure as in a recent paper for a resilient disk in an infinite baffle,[10] it can be shown that

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

T. Mellow and L. Kärkkäinen: Membrane in free space and baffle   2473

$$I_{\text{Fin}}(m,w,z) = -\sum_{p=0}^{P}\sum_{q=0}^{Q}\frac{1}{p!q!(ikz)^{2(p+q)+1}}\left(\frac{ka}{2}\right)^{p-m+(1/2)}$$
$$\times\left(\frac{kw}{2}\right)^{q}J_{p+m+(3/2)}(ka)J_{q}(kw)$$
$$\times\gamma(2p+2q+1,ikz). \tag{113}$$

c. Solution of the infinite integral. Similarly, it can be shown that[10]

$$I_{\text{Inf}}(m,w,z) = \left(\frac{2}{ka}\right)^{2m+1}\sum_{p=0}^{P}\sum_{q=0}^{Q}\frac{(-1)^{p+q}(2p+m+3/2)}{q!(k^2z^2+k^2a^2)^{q-(m/2)-(1/4)}}$$
$$\times (p+1)_{m+(1/2)}\Gamma(2p+2q+1)\left(\frac{kw}{2}\right)^{q}$$
$$\times J_{2p+m+(3/2)}(ka)J_{q}(kw)P_{2q-m-(3/2)}^{-2p-m-(3/2)}$$
$$\times\left(\frac{z}{\sqrt{z^2+a^2}}\right). \tag{114}$$

The normalized near-field pressure response of the membrane is shown in Fig. 16 for various values of ka. The calculations were performed using 50 digit precision with $P=M=10+ka$ and $Q=2P$.
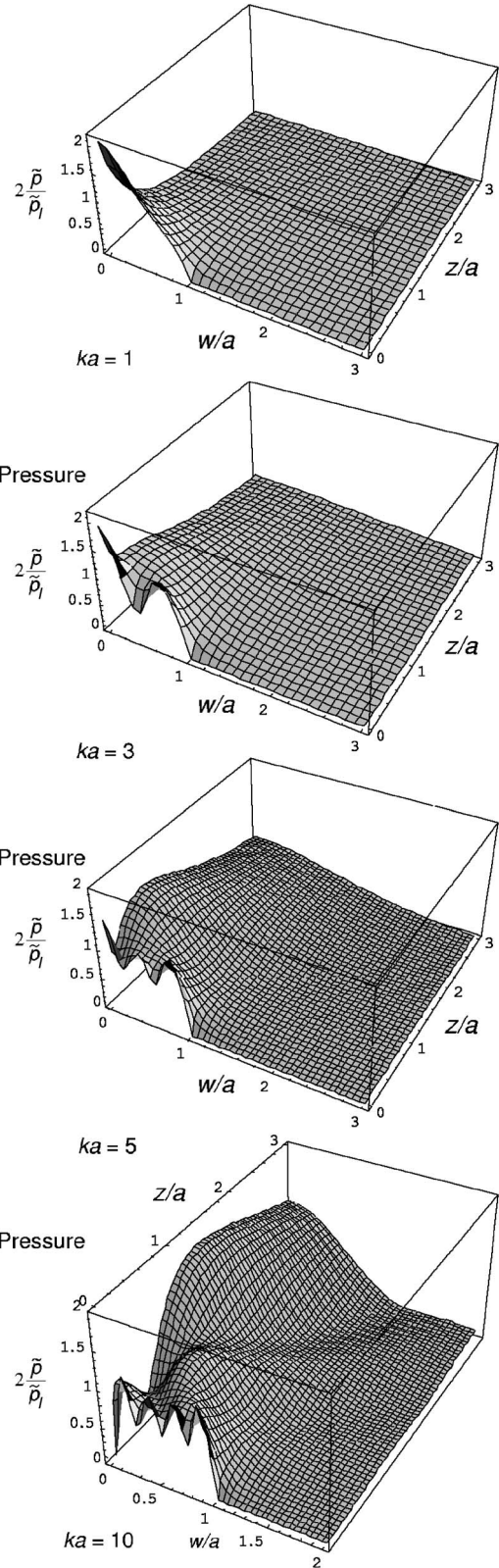
H. Far-field pressure response

The far-field pressure is derived using the same procedure as shown in Part I of Sec. II of a recent paper[10] to give the same result as Eq. (79), except that the directivity function $D(\theta)$ is now given by

$$D(\theta) = ka\sum_{m=0}^{M}\tau_m\Gamma\left(m+\frac{5}{2}\right)\left(\frac{2}{ka\sin\theta}\right)^{m+(3/2)}$$
$$\times J_{m+(3/2)}(ka\sin\theta), \tag{115}$$

which is the same as for the free space membrane except for the absence of the $\cos\theta$ term because the free space membrane is a dipole, whereas the baffled membrane is a monopole. For $\theta=0$ (i.e., on-axis), this simplifies to

$$D(0) = ka\sum_{m=0}^{M}\tau_m = \frac{ka}{Z_I}, \tag{116}$$

using Z_I from Eq. (106). Hence the on-axis pressure simplifies to

$$\tilde{p}(r,0) = i\frac{ka^2}{4r}\tilde{p}_I e^{-ikr}\frac{1}{Z_I} = i\omega\frac{E_P}{d}C_{\text{ED}}\frac{\tilde{e}_{\text{in}}}{2\pi rc}e^{-ikr}\frac{1}{Z_I}. \tag{117}$$

Taking into account that $\tilde{I}_{\text{in}}\approx i\omega C_{\text{ED}}\tilde{e}_{\text{in}}$, leads to

$$\tilde{p}(r,0)|_{b=a} \approx \frac{1}{Z_I}\cdot\frac{E_P}{d}\cdot\frac{\tilde{I}_{\text{in}}}{c}\cdot\frac{e^{-ikr}}{2\pi r}, \tag{118}$$

which, in the case of $Z_I=1$, is Walker's equation.[21] A better approximation is given by

$$Z_I \approx 1 + R_S/(2\rho c), \quad ka > 2, \tag{119}$$

which leads to what could be termed the "modified Walker's equation." The on-axis pressure response with damping is shown in Fig. 17, together with the efficiency and peak dis-

placement at $w=0$. In Fig. 18 is shown a comparison of the on-axis responses with and without damping (the level is reduced in order to accommodate the large undamped excursions), together with the high frequency approximation, where the SPL is given by: $\text{SPL}=20\log_{10}|\tilde{p}(r,0)/(20\times 10^{-6})|$. The normalized directivity function $20\log_{10}|D(\theta)|/|D(0)|$ is plotted in Fig. 19 for various values of ka.

IV. FINITE ELEMENT MODEL

Abaqus CAE v6.5-5 was used for the finite element analysis. In order to save processing power, only a quarter membrane was simulated. The analysis was carried out in two steps using a cylindrical coordinate system. In the first step, a boundary condition was applied to rim of the quarter membrane, which blocked axial or tangential displacement, but allowed radial displacement. Also, a boundary condition of zero tangential displacement was applied to its two straight edges. A pretension T was applied to the rim node set using the formula

$$\text{Force per node} = 0.5\pi aT/(\text{No. of nodes}). \tag{120}$$

The second step was a steady-state dynamics analysis in which an evenly distributed pressure of $\tilde{p}_I=1$ Pa was applied to the membrane surface. A limitation is that there is no obvious way to simulate the viscous acoustic resistance presented by a perforated screen or mesh. The elements used for the simulation are described in Table I.

The membrane was surrounded by a quarter-sphere of air with a radius of 0.5 m. The outer curved surface of this quarter-sphere was coated with a skin of infinite elements in order to prevent reflections. The air space was created as two separate front and rear segments, which both had their adjoining faces partitioned where they met the rim of the membrane. Each was tied to one of the surfaces of the membrane and, in the case of the unbaffled membrane, to each other beyond the rim. Hence, in order to simulate an infinite baffle, the coupling between the air segments was omitted so that the rigid boundary condition at the baffle was created auto-
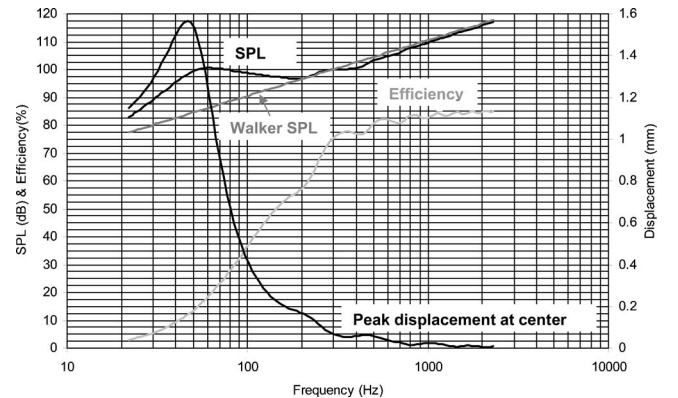
FIG. 17. On-axis pressure response of push-pull electrostatic loudspeaker in infinite baffle where $a=250$ mm, $h=12$ μm, T=700 N/m, $\rho_D=1400$ kg/m³, $e_{\text{in}}=2000$ V$_{\text{rms}}$, $E_P=3000$ V, $d=2$ mm, $r=1$ m, and $R_S=100$ Ns/m³.
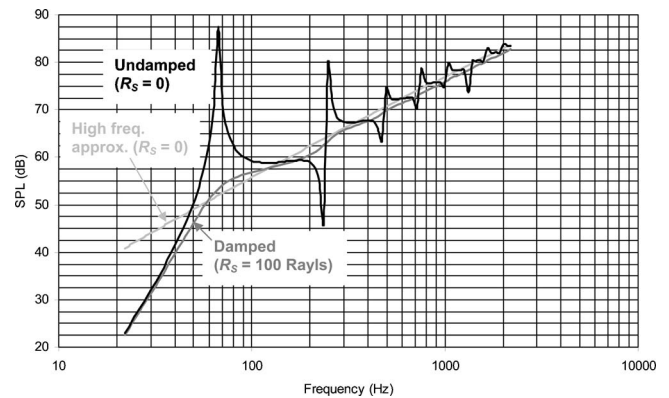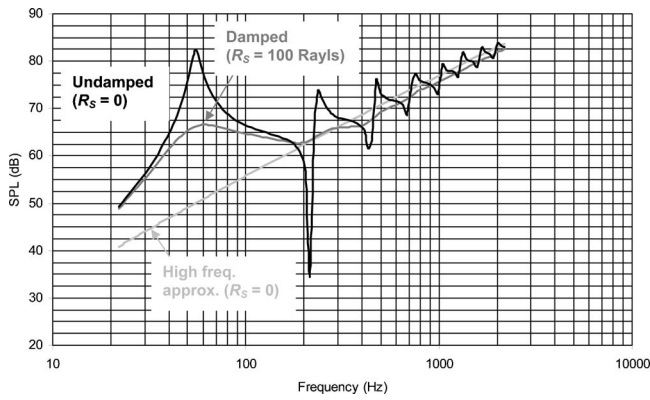
FIG. 18. On-axis pressure response with same parameters as in Fig. 18, except $e_{in}=40$ V$_{rms}$, and with $R_S=0$ (undamped).

matically without having to create a separate baffle surface. This also allowed the same mesh to be used for both simulations.

The outer curved edges of the air space were seeded with a node density of 80 nodes/m (or 6 nodes/wavelength in free space at 4.6 kHz). The edges of the membrane were seeded with higher density of 400 nodes/m (or 6 nodes/wavelength in the membrane at 4.7 kHz), as were also the adjacent edges of the air space to which its edges were tied. The remaining straight edges of the air space were left unseeded. Free meshing was used throughout. The total number of air elements was 588 757 with minimum and maximum angles of 7.5° and 143.4°, respectively. The number of membrane elements was 4824 with minimum and maximum angles of 29° and 115°, respectively. The finite size of the air space produced some errors below 170 Hz, so the analysis at these frequencies was performed using a 2.5 m radius quarter-sphere with a coarser mesh density to save memory and processing time. The on-axis responses without and with a baffle are shown in Figs. 20 and 21, respectively.

V. DISCUSSION

At low to medium frequencies, the analytical calculations and FEM results show excellent agreement, as can be seen in Figs. 20 and 21. However, above 1 kHz some divergence occurs. This is in no doubt partly due to the discretization of the finite element model, which has a density of 6 elements per wavelength at the highest frequency shown. There could also be some differences due to the fact that the FEM response was calculated at a distance 0.5 m from the source and then adjusted to 1 m by subtracting 6 dB, whereas the analytical result was calculated using the far-field asymptotic expression. The analytical calculations for the free space membrane without damping show good agreement with Streng's results[5] too.

It is interesting to see how the free space radiation impedance in Fig. 5 is an almost monotonic function when compared with that of a rigid disc. In this respect, it is more like that of a resilient disc in free space [13] (or "freely suspended disc"),[14] which is perhaps not surprising, considering that the membrane is essentially a more or less evenly distributed pressure source (i.e., transparent) rather than velocity one (i.e., opaque). As can be seen from the near-field pres-
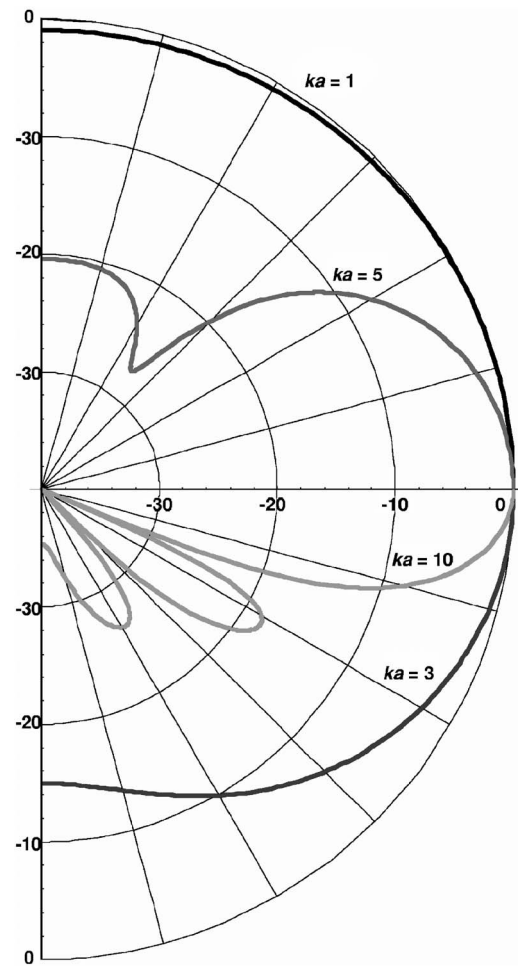
FIG. 19. Normalized far-field directivity function in infinite baffle with same parameters as in Fig. 17.

sure plots of Fig. 8, the surface pressure becomes more evenly distributed at higher values of ka. However, one important difference between a membrane and a resilient disc is that the membrane is clamped at its outer edge whereas the resilient disc is free and therefore not realizable.[13] Interestingly, the real part of the membrane's radiation impedance at low frequencies is greater than that of the rigid disk, possibly due to the fact that its perimeter is clamped so that most of the sound is radiated from the center, with the remainder acting as a semibaffle to mitigate the acoustic "short-circuit" between the front and rear surfaces.

The baffled radiation impedance in Fig. 14 is also a fairly smooth function, compared to that of a rigid disk, and is therefore closer to the characteristic of a resilient disk in an infinite baffle[10] (or plane wave diffracted through a circu-

TABLE I. Summary of the finite element model.

Material	Section type	Element type
Membrane	Membrane	M3D3 Tri, 3 Node Triangular Membrane
Air	Solid Homogeneous	AC3D4 Tet, 4 Node Linear Acoustic Tetrahedron
Air	Solid Homogeneous.	ACIN3D3 Tri, 3 Node Acoustic Infinite
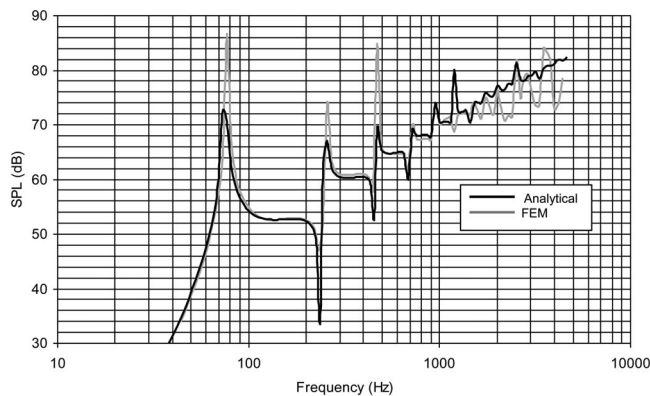
FIG. 20. On-axis pressure response in free space without damping: comparison of analytical and FEM models, where $a=125$ mm, $h=12$ μm, $T=100$ N/m, $\rho_D=1667$ kg/m^3, $p_I=1$ Pa, $r=1$ m, and $z_S=0$.

FIG. 21. On-axis pressure response in infinite baffle without damping: comparison of analytical and FEM models, where $a=125$ mm, $h=12$ μm, $T=100$ N/m, $\rho_D=1667$ kg/m^3, $p_I=1$ Pa, $r=1$ m, and $z_S=0$.

lar aperture)[6]. As can be seen from the near-field pressure plots of Fig. 16, the surface pressure eventually becomes more evenly distributed at higher values of ka and so converges towards the flat pressure distribution of a resilient disk.

It can be seen from Figs. 4 and 13 that the motional impedance in the "piston range" above 440 Hz (or $ka>2$) is mainly resistive, as is also the radiation impedance shown in Figs. 5 and 14. In this region, as the commonly used term "piston range" suggests, the diaphragm effectively radiates sound power into the surrounding space and it is here that the efficiency (see Figs. 9 and 17) is highest. The rising on-axis response in this region appears to be fairly consistent with that predicted by Walker's equation, a fairly detailed discussion of which is provided by Baxendall[1] (edited by Borwick). This equation generally describes the free space on-axis pressure response (with some damping) between the fundamental resonance and the point at which the diaphragm inertia, together with the acoustic (mainly radiation) resistance, starts to take control. If the electrode structure were to be partitioned into concentric rings, as described by Walker,[21] the on-axis response in the piston range could, in theory, be equalized and it would be interesting to see this included in the model. Also, the polar responses of Figs. 11 and 19 could be widened without losing power. The effect of the baffle is to equalize the on-axis response in the lower frequency range ($ka>2$) so that, with electrode partitioning, it could be fairly flat all the way from the fundamental resonance to the high frequency roll-off, at which the diaphragm inertia begins to dominate.

The motional impedance has a minimum value at the fundamental resonance, which is at 67 Hz in free space (see Fig. 4) and 53 Hz in a baffle (see Fig. 13). Here, the efficiency is lowest because most of the power is being dissipated in the acoustic damping resistance. (This behavior is in contrast with that of an electrodynamic loudspeaker, which tends to be most efficient at the suspension resonance and then becomes less efficient with increasing frequency.) Not surprisingly, adding a baffle greatly improves the efficiency in this region. Below the fundamental resonance, the slope of
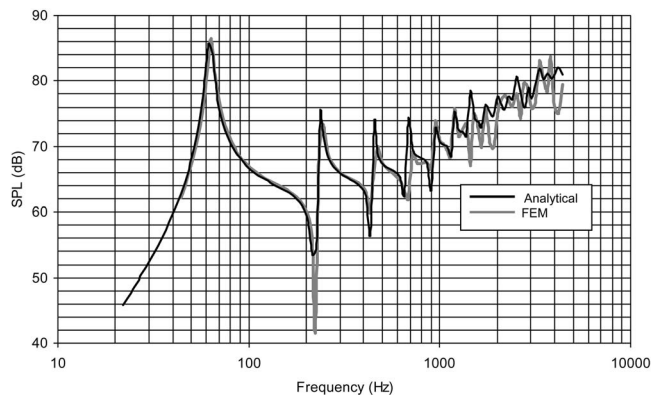
the on-axis response is 18 dB/octave in free space and 12 dB/octave in a baffle.

VI. CONCLUSIONS

A more direct method has been presented for calculating the radiation characteristics of a circular membrane in free space than previously available. Furthermore, the method has been extended to allow for the inclusion of an infinite baffle. In particular, equations have been derived for the motional impedance and diaphragm impedance, as well as expansions for the pressure field. The final equations for calculating the power series coefficients of the free space and baffled membranes are remarkably similar in form. Most of the terms are the same, but are just arranged differently. However, the pressure field is much easier to calculate for the baffled membrane when the distance to the observation point is greater than its radius.

It has also been shown how this model can be used as a benchmark for simulation using tools such as FEM. It is not intended, though, in this analytical versus FEM comparison, to select a "winner." On the contrary, the two approaches appear to be somewhat complementary. Whereas FEM is better suited to complicated geometries, analytical simulations tend to be limited to simply geometry, typically axisymmetric (i.e., two-dimensional). On the other hand, the analytical approach has distinct advantages at the frequency extremes and in deriving far-field asymptotic expressions. Furthermore, describing a model with equations provides us with a greater insight into the actual physics of the problem.

[1]J. Borwick, *Loudspeaker and Headphone Handbook*, 3rd ed. (Focal, Oxford, 2001), pp. 108–195.
[2]J. Zemanek, "Beam behaviour within the nearfield of a vibrating piston," J. Acoust. Soc. Am. **49**(1), 181–191 (1971).
[3]A. Nisbett, *The Use of Microphones*, 4th ed. (Focal, Oxford, 1993), p. 44.
[4]J. H. Streng, "Calculation of the surface pressure on a vibrating circular stretched membrane in free space," J. Acoust. Soc. Am. **82**(2), 679–686 (1987).

[5] J. H. Streng, "Sound radiation from a circular stretched membrane in free space," J. Audio Eng. Soc. **37**(3), 107–118 (1989).

[6] C. J. Bouwkamp, "Theoretical and numerical treatment of diffraction through a circular aperture," Indian J. Pure Appl. Phys. **AP18**(2), 152–176 (1970).

[7] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 320, 365, 389, 390.

[8] R. D. Spence, "The diffraction of sound by circular disks and apertures," J. Acoust. Soc. Am. **20**(4), 380–386 (1948).

[9] R. D. Spence, "A note on the Kirchhoff approximation in diffraction theory," J. Acoust. Soc. Am. **21**(2), 98–100 (1949).

[10] T. J. Mellow, "On the sound field of a resilient disk in an infinite baffle," J. Acoust. Soc. Am. **120**(1), 90–101 (2006).

[11] H. Suzuki and J. Tichy, "Sound radiation from an elastically supported circular plate," J. Acoust. Soc. Am. **65**(1), 106–111 (1979).

[12] T. J. Mellow and L. M. Kärkkäinen, "On the sound field of a membrane in an infinite baffle," on the CD ROM: Audio Engineering Society Convention Papers, 118th Convention, Barcelona, 28–31 May 2005, available from Audio Engineering Society Inc., 60 East 42nd Street, Room 2520, New York, NY 10165–2520.

[13] T. J. Mellow and L. M. Kärkkäinen, "On the sound field of an oscillating disk in an open and closed circular baffle," J. Acoust. Soc. Am. **118**(3), 1311–1325 (2005).

[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed., edited by A. Jeffrey (Academic, New York, 2000), p. 658, Eq. (6.521.1), p. 671, Eq. (6.567.1), p. 668, Eq. (6.561.5), p. 342, Eq. (3.381.1), p. 930, Eq. (8.532.1), p. 980, Eq. (8.930.1)–(8.930.7), p. 342, Eq. (3.381.4), p. 891, Eq. (8.356.3), p. 691, Eq. (6.621.1).

[15] S. Wolfram, *The Mathematica Book*, 5th ed. (Wolfram Media, Champaign, IL, 2003). Symbolic computation by MATHEMATICA.

[16] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964), pp. 487, Eq. (11.4.41).

[17] L. L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1993), pp. 101–105, 118–128.

[18] M. R. Stinson, "The propagation of plane sound waves in narrow and wide circular tubes, and generalization to uniform tubes of arbitrary cross-sectional shape," J. Acoust. Soc. Am. **89**(2), 1311–1325 (2005).

[19] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge University Press, London, 1944), pp. 141, Sec. 5.22, Eq. (5).

[20] A. O. Williams, "Acoustic field of a circular plane piston," J. Acoust. Soc. Am. **36**(12), 2408–2410 (1964).

[21] P. J. Walker, "New Developments in Electrostatic Loudspeakers," J. Audio Eng. Soc. **28**(11), 795–799 (1980).

Multiload procedure to measure the acoustic scattering matrix of a duct discontinuity for higher order mode propagation conditions

Azzedine Sitel, Jean-Michel Ville,[a] and Felix Foucart
Laboratoire Roberval FRE UTC-CNRS No. 2833, Université de Technologie de Compiègne,
BP 20529 F60205, Compiègne Cedex, France

A procedure for measuring the acoustic scattering matrix coefficients of a duct discontinuity for higher order acoustic duct mode propagation conditions is described and tested. The technique requires measurement of pressure waves per mode coming in and out of the discontinuity. Assuming N cut-on modes, the $(2N)^2$ scattering matrix coefficients are determined after repeating the experiment for N linearly independent pressure distributions for at least two load configurations. Experiments were conducted for a straight duct and a reactive chamber. A good agreement was found between experiment and theory except near cut-off frequencies. The overdetermination method based on four loads was shown to improve the results. An analytical simulation of the experiment was developed to compute the influence on the $[S]$ calculation of an error in temperature and total modal pressure assumed to be representative of a real measurement situation. This simulation with a discussion explains the discrepancies between experiment and theory. The test with the chamber shows that the load method fails as expected in determining the coefficients associated to the wave coming in the discontinuity from the open end side because of the property of the middle duct to filter modes making the results very sensitive to uncertainties. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354040]

I. INTRODUCTION

Sound propagation in duct systems is a problem of considerable practical interest for automotive, building, and aeronautic industries. A pipe system is usually considered as a network of straight ducts coupled together via various passive elements such as mufflers, bends, T-sections, etc. In order to analyze and optimize sound transmission, reflection, and radiation through and from those basic elements a formalism based on transfer matrix method[1,2] has been developed and most often used by manufacturers to design, for example, exhaust lines and air conditioning circuits. The scattering matrix formalism which uses the traveling wave amplitudes as state variables has been shown to be more attractive than transfer or mobility matrices since it reflects the fundamental waveguide nature of the problem allowing one to develop, for example, a general proposal for analyzing acoustic two-port networks[3,4] or to deduce acoustic energy dissipation or production in a multiport.[5]

Over the years a number of papers have been published on techniques based on two-loads and two source methods to perform measurements of the scattering matrix for plane wave propagation condition.[6,7]

Until now no specific measurement procedure which can be applied in higher order mode propagation conditions has been achieved. Only investigations on experimental methods to measure reflection and transmission matrices of discontinuities in a duct have been conducted.[8,9]

The aim of the paper is to present and test a method to measure the scattering matrix $[S]$ of a discontinuity for higher order mode propagation conditions. Unlike the reflection or transmission matrices, the scattering matrix of size $2N \times 2N$ (N being the number of cut-on modes) gives a more complete description of the discontinuity. Indeed it is independent of the open end duct conditions and can then be integrated in computer programs used for the conception of complex duct networks.

In Sec. II, after presenting the definition of the scattering matrix, the theory which the method is based on is described. In Sec. III, the experimental facility, the two duct discontinuities tested, and the procedure which leads to the determination of the scattering matrix are detailed. In the Sec. IV the experimental results are compared with theory for both duct configurations. Moreover an analytical simulation of the experiment is performed to study the sensibility of the $[S]$ coefficients to some parameters of the experiment such as temperature, microphone locations, and modal coefficients and used to explain some discrepancies observed between experiment and theory.

II. THEORETICAL BASIS

A. Definition of the scattering matrix $[S]$ of a discontinuity

The scattering matrix $[S]$ of a discontinuity (in grey in Fig. 1) located between the axial coordinates z_L on side I and z_R on side II (L and R mean, respectively, the left side and right side of the discontinuity) and linked to two waveguides is a linear relationship between the incoming pressure wave
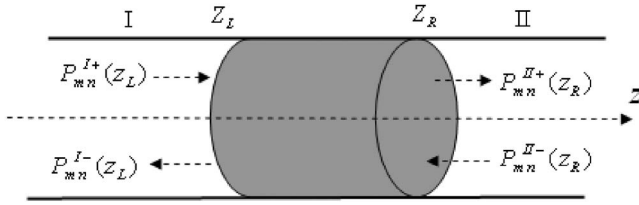
FIG. 1. Waves coming in and out of the discontinuity in gray from sides I and II.

vector $\{P^{in}\}_{2N}=\{P^{I+}_{mn}(z_L), P^{II-}_{mn}(z_R)\}^t_N$ and the outcoming pressure wave vector $\{P^{out}\}_{2N}=\{P^{I-}_{mn}(z_L), P^{II+}_{mn}(z_R)\}^t_N$ where N is the number of modes in both cross sections:[10]

$$\{P^{out}\}_{2N} = [S]_{2N\times 2N} \times \{P^{in}\}_{2N} \quad \text{with} \tag{1}$$

$$[S] = \begin{bmatrix} [S^{11}]_{N\times N} & [S^{12}]_{N\times N} \\ [S^{21}]_{N\times N} & [S^{22}]_{N\times N} \end{bmatrix}_{2N\times 2N}.$$

This matrix, which depicts only the discontinuity, is independent of the upstream and downstream acoustic conditions and is filled by $(2N)^2$ coefficients:

$$S^{kl}_{mn,pq}, \quad k,l = 1,2 \quad \text{and} \quad m,n,p,q = 1,2,\dots,N.$$

The physical meaning of each elementary matrix is as follows: $[S^{11}_{mn,pq}]_{N\times N}$ is the reflection of the wave coming in the element from the left side; $[S^{21}_{mn,pq}]_{N\times N}$ is the transmission of the wave coming in the element from the left side; $[S^{22}_{mn,pq}]_{N\times N}$ is the reflection of the wave coming in the element from the right side; $[S^{12}_{mn,pq}]_{N\times N}$ is the transmission of the wave coming in the element from the right side.

B. Theory of the procedure to measure the scattering matrix [S] of a discontinuity

In no flow conditions, the fluid is assumed to be ideal and linear acoustic theory to be valid. As the discontinuity under test is connected to and between two hard wall cylindrical ducts with radius a (main duct) at axial coordinates z_L and z_R (Fig. 1), the acoustic pressure distribution in the main duct is written[11] in cylindrical coordinates (r, Θ, z) as

$$p(r,\theta,z,t) = \sum_{m=-\infty}^{+\infty} \sum_{n=0}^{\infty} P_{mn}(z)\Psi^a_{mn}(r,\theta)e^{-i\omega t}, \tag{2}$$

where $\Psi^a_{mn}(r,\theta)=J_m(\chi_{mn}/a)e^{im\theta}$ are the eigenfunctions; the integers m and n are, respectively, the angular wave number and the radial number ($m=0$, $n=0$ is the plane wave); J_m is the Bessel function of the first kind of order m; χ_{mn}/a is the nth root satisfying the radial hard wall boundary condition on the wall of the main duct $J'_m(\chi_{mn}/a)=0$. The total modal pressure coefficients $P_{mn}(z)$ in the cross sections located at axial coordinates z_L and z_R are given by the following relationship:

$$P_{mn}(z) = [P^+_{mn}(z) + P^-_{mn}(z)], \tag{3}$$

where pressure waves traveling in the positive and negative z directions are depicted, respectively, in regions I and II by $P^{I\pm}_{mn}(z)=P^{I\pm}_{mn}(0)e^{\pm ik_{mn}z}$ and $P^{II\pm}_{mn}(z)=P^{II\pm}_{mn}(0)e^{\pm ik_{mn}z}$. k_{mn}

$= \sqrt{k^2-(\chi_{mn}/a)^2}$ is the axial wave number of mode (m,n) in the main duct, $k=2\pi f/c_0$, f the frequency, and c_0 the speed of sound in air.

The $(2N)^2$ coefficients of $[S]$ are solutions of the following system deduced from Eq. (1):

$$[S^{kl}_{mn,pq}]_{2N\times 2N} = [P^{out}]_{2N\times 2N} \times [P^{in}]^{-1}_{2N\times 2N}. \tag{4}$$

To fill the $2N$ columns of $[P^{out}]_{2N\times 2N}$ and $[P^{in}]_{2N\times 2N}$, $2N$ linearly independent modal pressure distributions have to be produced and measured by the experiment in both cross sections located at z_L and z_R and incident and reflected modal pressure waves separated.

III. THE EXPERIMENTAL SETUP AND DATA PROCESSING

The measurement of $[S]$ was carried out with the duct facility built under a grant with E.E.C. during the Ducat program,[12] which was already used to measure reflection and transmission matrices.[9] The pressure distributions measured in two closed cross sections before and after the duct under test by two pairs of microphones are Fourier Lommel's transform[13] then the incident and reflected modes are separated.[9] If N is now the number of cut-on modes, a "selective" computation of the $2N\times 2N$ coefficients of $[S]$ is then performed after a process based on the generation of N independent incident fields for two- or six-load cases depending on the direct or overdetermination procedure is achieved. The load method was chosen instead of the source[7] one because of practical considerations.

A. Hardware

The facility is installed in the anechoic chamber of the University of Technology of Compiegne. The equipment is made of 0.5-m-long duct components with 0.01-m-thick steel wall which all have a 0.148-m internal diameter except of course part of the test duct element. From the left to the right of the schema (Fig. 2) are shown the following elements:

(1) Absorbent element to avoid reflection upstream.
(2) The source section with three acoustic drivers flushed mounted in a z line. The axial distance between two drivers is 0.15 m. All this section can rotate over 360°.
(3) The measurement duct element I which can rotate over 360° is supporting a boom traversing on the radius with a B&K sound intensity probe attached and directed toward the z axis. The distance between the 1/4 in. microphones is 0.054 m according to well-known conditions.[8,13,14]
(4) The 0.5-m-long duct element which will be part of the test configuration.
(5) The measurement duct element II is identical to the duct element I.

Noise is radiated in the anechoic chamber through an unflanged inlet duct which constitutes the load. The different load conditions will be created by changing only the length of this duct. Angular and radial displacements are provided by step by step motors. A working station automatically op-

FIG. 2. Experimental setup.

1- Absorbing foam.
2- Source element.
3- Measurement element I.
4- Test element.

5- Measurement element II.
6- Load.
7- Sound intensity probe.
8- Sources.
9- Step by step motors.

FIG. 3. Flow chart of the data collection in both closed duct sections and separation procedure.
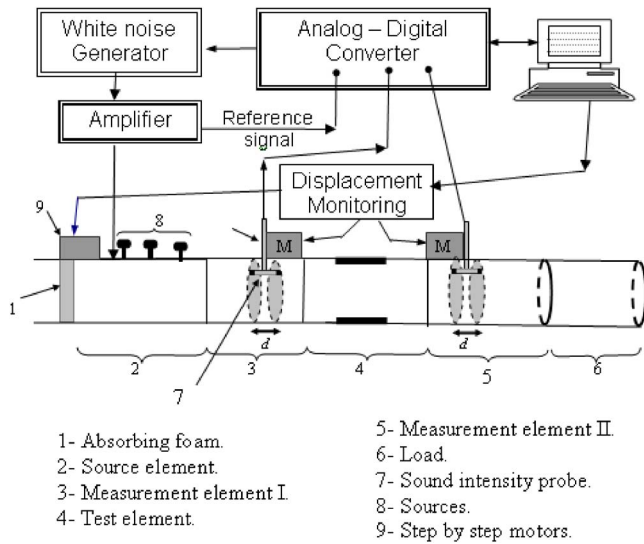
erates the rotation of the source section, the choice of the axial source, and the displacements of the probe in the r and Θ directions. It also supplies the noise generation to sources and the acquisition through an analog-digital converter of the signals issued from the four microphones.

B. The data processing

The experimental procedure is divided into the following steps:

(1) For each load and source configuration:

   (a) Collect the frequency spectra of the total modal coefficients in two closed cross sections located in both measurement duct elements.

   (b) Separate the incident and reflected modal pressure vectors.

(2) Postprocess the incident and reflected modal pressure vectors on both sides of the test duct element for the N source configurations and two or several loads to get the in and out wave matrices in z_L and z_R.

(3) Compute the scattering matrix.

1. Data collection and reduction

During the first step described in Fig. 3 the acoustic driver is driven with a white noise signal that is band limited to 0–2700 Hz. For this duct radius and frequency range the total adimensional wave number ka is lower than 3.8. Therefore only $N \leqslant 5$ modes (0,0); (± 1,0); (± 2,0) are cut-on. The transfer functions between the amplifier and microphone signals provide, after calibration, the amplitude and phase of the local total acoustic pressure normalized by the level of the amplifier. The origin of the z axis is then given by the reference phase that is in the source z axis position. A 240-point discretization of the acoustic pressure field measured in each of the two pairs of cross-section areas is achieved by rotating the measurement duct section at 16 positions equally spaced by 22.5° and displacing the probe to 15 radial positions.
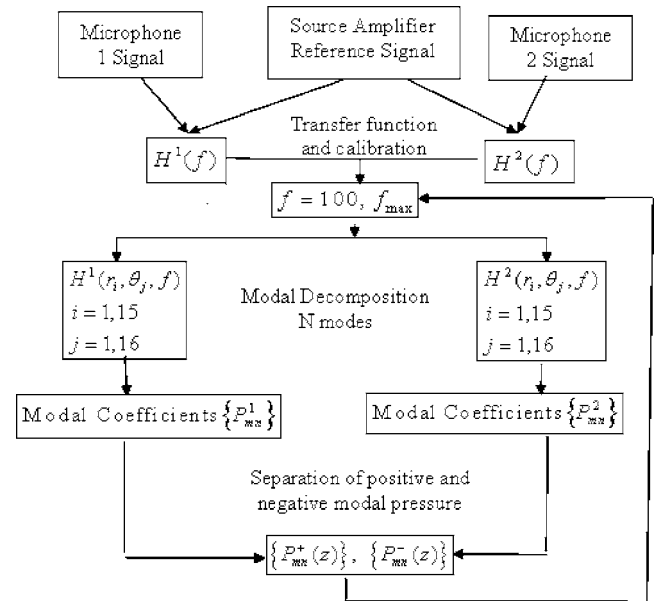
Then modal decomposition by a Fourier-Lommel's transform[8,13] for $m = -7$ to $+7$ and $n = 0$ to 3 is achieved leading to the cut-on or cut-off complex modal pressure coefficients in both pairs of cross-section areas. The incident and reflected modal vectors are then separated and computed at z_L and z_R.

All this procedure is repeated for all source configurations and loads. To compute the 100 coefficients of the matrix $[S]$, at least 10 experiments (5 source configurations and 2 loads) have to be performed. The 5 source configurations and loads have been chosen[8] to ensure generation of linearly independent input modal vectors. The coordinates of the source and the characteristics of the loads used are given in Table I.

2. Data postprocessing (Fig. 4)

At the end of the first step the modal coefficient vectors $\{P_{mn}^{I+}(z_L)\}\{P_{mn}^{II+}(z_R)\}$ of the positive waves and $\{P_{mn}^{I-}(z_L)\}$ $\times\{P_{mn}^{II-}(z_R)\}$ of the negative waves are available at each side of the discontinuity for each source and load configurations and the matrices $[P^{in}]_{2N \times 2N}$ and $[P^{out}]_{2N \times 2N}$ are filled. The scattering matrix is then deduced from the relationship Eq. (4) during the second step as shown in Fig. 4.

Equation (4) can be solved in two ways:

(1) With the "selective" method, the rank of $[S]$ is linked to the number of cut-on modes and will then depend upon the frequency. Only the propagating modes are taken into account assuming that the influence of the cut-off modes can be neglected.

(2) With the "global" method, the rank of $[S]$ is constant being independent of the frequency and fixed in relation to the higher limit of the frequency band studied, i.e., 2700 Hz, $N = 5$ and the rank of $[S] = 100$ in our experiment. The overall modal basis is considered including the nonpropagating modes.

TABLE I. Coordinates of the source and length of the load configurations.

Configuration No.	Angular position $\theta_S^i(°)$ of the source	Axial position z_S^j (cm) of the source	Length of the load E (cm)
1	0	0	0
2	120	+15	3
3	240	−15	6
4	8.5	0	9
5	128.5	+15	12
6	248.5	−15	18
7	17	0	0
8	137	+15	3
9	257	−15	6
10	25.5	0	9
11	145.5	+15	12
12	265.5	−15	18
13	34	0	0
14	154	+15	3
15	274	−15	6
16	42.5	0	9
17	162.5	+15	12
18	282.5	−15	18
19	51	0	0
20	171	+15	3
21	291	−15	6
22	59.5	0	9
23	179.5	+15	12
24	299.5	−15	18
25	68	0	0
26	188	+15	3
27	308	−15	6
28	76.5	0	9
29	196.5	+15	12
30	316.5	−15	18

FIG. 4. Flow chart of the scattering matrix coefficients calculation.

C. The configurations tested

The axial positions z_R and z_L defined in Fig. 1 correspond, respectively, to the axial coordinates of the two microphones located on each side of the duct under test. Then the duct element for which the scattering matrix is measured is 1 m long including the 0.5-m-long test duct and parts of the measurement ducts. The two discontinuities tested are symmetrical in relation to the duct axis and to the axial centerline. The first configuration is a 0.5-m-long element duct which is identical to the other elements. It has been chosen to become a reference test because its theoretical scattering matrix is well known (Appendix B) and also because the experimental conditions are more favorable with respect to the stationary wave problem. The second configuration is a reactive type chamber shown in Fig. 5 which introduces more complicated measurement conditions. It is made with the reference 0.5-m-long straight duct in which a 5.5-cm-diam axisymmetric cylindrical duct is mounted.

IV. ANALYSIS OF THE EXPERIMENTAL RESULTS

To analyze the experimental results several theoretical developments have been achieved:

Later on in this paper (Sec. V C 2), the experimental results of the calculation of a $[S]$ coefficient (Fig. 19) deduced from both techniques will be compared and show they provide nearly the same result. Therefore the "selective" method has been chosen to solve Eq. (4). It is well known that this direct method based upon the inversion of the $[P^{in}]_{2N \times 2N}$ matrix [Eq. (4)] leads to numerical problems. To avoid this difficulty an overdetermination technique[15,16] often used to solve a system where the number of equations $L \times N$ ($L > 2$) is higher than $2N$ the number of unknowns was developed. The scattering matrix $[S]$ is then now deduced from the relationship:[17]

$$[S]_{2N \times 2N}^T = [[^*[P^{in}]_{2N \times (L \times N)}$$
$$\times [P^{in}]_{2N \times (L \times N)}^T]^{-1}$$
$$\times {}^*[P^{in}]_{2N \times (L \times N)}] \cdot [P^{out}]_{2N \times (L \times N)}^T. \quad (5)$$

The columns of the rectangular matrices $[P^{in}]_{2N \times (L \times N)}$ and $[P^{out}]_{2N \times (L \times N)}$ are filled by repeating N experiments associated to N source configurations for $L = 4$ or 6 load configurations ensuring[8] again that the $L \times N$ pressure distributions are linearly independent.

FIG. 5. The reactive muffler located between z_L and z_R.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Sitel *et al.*: Acoustic scattering matrix measurement   2481

(1) A theoretical calculation of $[S]$ for both configurations which the experimental results are compared with.

(2) An analytical simulation of the experiment and an error calculation on the determination of $\{P_{mn}^{I+}(z_L)\}$ and $\{P_{mn}^{I-}(z_L)\}$ by the separation procedure to discuss the discrepancies between experiment and theory.

Among the $(2N)^2 = 100$ coefficients $(N=5)$ of $[S]$ only typical results for the plane wave and for higher order modes $m=1$ and 2 are presented emphasizing the symmetry properties of both discontinuities.

A. Influence of errors

The resulting error on the calculation of $[S]$ will depend on two things: (1) the errors in the input measured data, i.e., temperature, transfer functions, and position of the microphones; and (2) the sensivity to errors in the input data of the calculation formulas, i.e., induct modal decomposition, incident and reflected waves separation, and calculation of $[S]$.

During a previous study[14] on the influence of errors on the two-microphone separation method assuming plane wave propagation condition, some practical conclusions about the overall length of the duct, the reflective conditions on the source and open end sides were drawn in order to keep error in the transfer function estimate at 1% in the magnitude and 0.6° in phase. Also as a result of this study, a condition on the distance d between microphones which has to be respected in order to avoid a very large sensitivity to errors in input data during the separation procedure was deduced. In our experiment this distance d was chosen equal to 5.4 cm in order to respect this condition leading to an upper frequency limit of 2700 Hz (Sec. III A). For higher order mode propagation conditions, the sensitivity to the modal decomposition technique in the microphone location and signal to noise ratio was already discussed during previous works.[18] Assuming the technique to be a generalization of this developed for plane wave case[8,19] the optimal distance of separation was extended to higher order mode propagation conditions.

The calculation of error presented in this paper based upon a direct simulation of the experiment described in the following begins by the separation procedure and includes the calculation of $[S]$, then excluding the sensitivity to errors in the input data to the modal decomposition stage which is therefore considered as a known input data to the simulation program as shown in Fig. 6. Moreover, an analytical calculation of the error produced by the separation process based on a first-order Taylor expansion of the calculation formulas is presented in Appendix D. Results of this study will emphasize for higher order mode propagation condition the high sensitivity of this step to errors in input data such as temperature and total modal pressure when the frequency becomes very close to the cut-off mode frequency while for the plane wave the sensitivity to the error on temperature can be neglected compared to the error on total modal pressure.[20]

Another effect to be considered is the influence from the nonpropagating mode on the measurement of $[S]$. According to a study which discusses this matter for the plane wave case[14] the microphones should not be placed closer than about 10 mm from the discontinuity or a sample to avoid the
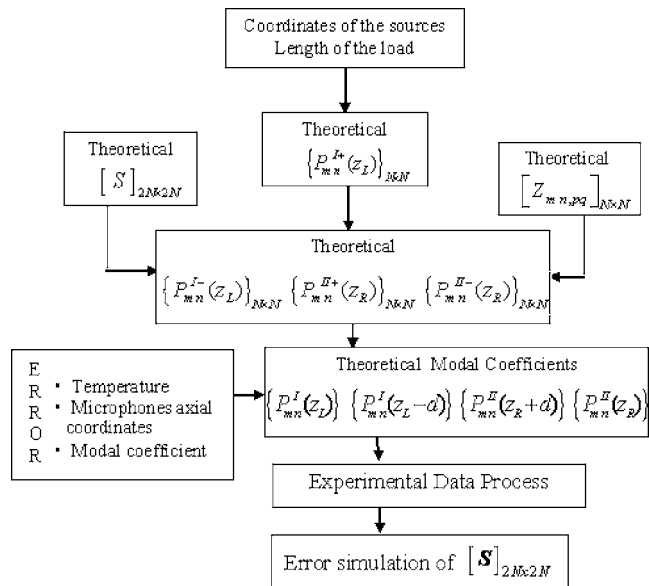
FIG. 6. Flow chart of the error estimation by a simulation process of the experiment.

effect of near field. For higher order mode propagation condition when the frequency approaches the cut-off frequency the influence of the evanescent mode increases as the "damping" length becomes long. In the present study the modal basis which the decomposition technique is calculated with, includes all modes up to $m = \pm 7$, $n = 3$ whatever the frequency is (Sec. III B 1) avoiding the influence of the evanescent modes up to the separation step included in the frequency band studied. Another problem is to include in the calculation of the matrix $[S]$ the conversion coefficients in reflection and transmission from a cut-on mode to a cut-off mode which assumes the modal basis used in Eq. (4) is extended to the nonpropagating modes. It will be shown in Sec. V C 2 that for an axisymmetric discontinuity in the frequency domain down to the cutoff frequency of the $(0,1)$ and as the distance between the discontinuity and the measurement cross section is far enough, the influence of the evanescent modes can be neglected. This assumption was already verified in previous studies where reflection and transmission matrices were measured.[8,9]

Then the direct simulation program has provided the calculation of error produced through the matrix scattering formulas for a 5° error on temperature and (1°, 5%) error on, respectively, the phase and the amplitude of the total modal pressure coefficients measured in both cross section values assumed to be representative of a real measurement situation.

B. Theoretical simulation of the experiment

1. Theory of the simulation

Figure 7 shows the duct arrangement used to simulate the experiment. It is made of the duct element located between z_L and z_R and defined by its theoretical scattering matrix $[S]$. On the left side of this duct element a semi-infinite duct is supporting the source section and on the right side a finite length duct is radiating outside through an infinite baffled open end. The total pressures in the four cross sections where the microphones are located are then calculated
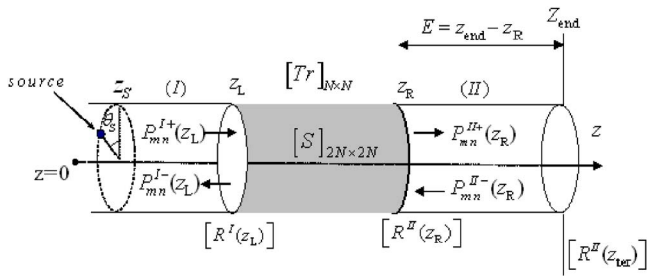
FIG. 7. Duct configuration of the simulation process.

Experiment (2 loads : - - -) ; Experiment (6 loads : ———)

Theory : — — —

FIG. 8. Comparison with theory for the straight duct configuration of $|S_{00,00}^{2,1}|$, the experimental transmission coefficient vs ka of the plane wave mode traveling from the left to the right side deduced from the direct and overdetermination methods.

analytically for each source and load configuration listed in Table I. To perform this simulation the following analytical calculations have to be carried out:

(1) The scattering matrix $[S]$ (Appendices A and B).
(2) The impedance matrix $[Z_{mn,pq}]_{N \times N}$ of the open end reflection.[19]
(3) The incident pressure matrix in the semi-infinite duct $\{P_{mn}^{I+}(z_L)\}_{N \times N}$ deduced for each of the N source configurations (Table I) assuming a point source defined by its amplitude $A_s = 1$ and phase $\varphi_s = 0$ and located at (z_s, θ_s, a).
(4) The modal pressure coefficients in both duct cross-section areas for all sources and loads configurations given in Table I (Appendix C).

2. The simulation procedure

The flow chart of the simulation procedure is described in Fig. 6. Its validation was carried out by introducing the theoretical values of $[S]$ and verifying that when no errors are assumed the procedure results in the same matrix $[S]$. The sensitivity of $[S]$ to errors on the following parameters were then studied by introducing errors on the temperature, modulus and phase of the modal pressure coefficients, and axial positions of the microphones.

C. Experimental results

1. Straight duct

Influence of the number of loads on the experimental results. The advantage of the overdetermination procedure over the direct method is clear in Fig. 8 where the variations of the transmission coefficient of the plane wave $S_{00,00}^{1,2}$ vs ka for two and six loads are compared with theory. All "instabilities" in the high frequency domain have been avoided when the overdetermination method is performed. As a consequence, all other results presented in the paper were carried out with the overdetermination method with six loads.

Analysis of some $[S]$ coefficients. The modulus of the transmission coefficients $[S_{00,00}^{1,2}; S_{00,00}^{2,1}]$ and reflection ones $[S_{00,00}^{1,1}; S_{00,00}^{2,2}]$ for the plane wave $(0,0)$ coming in the discontinuity from the left and right sides respectively, are plotted versus ka up to 3.8 and compared to the analytical results in Fig. 9. In Fig. 10 the same results are presented for the higher order mode case $(-1,0)$. At once the three following comments have to be made:
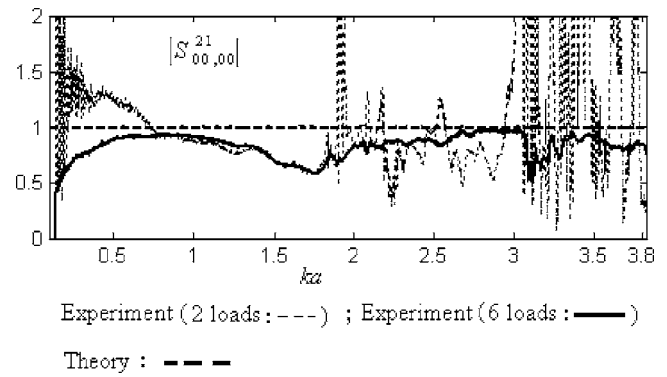
(1) The experimental curves $S_{m0,m0}^{2,1}$, $S_{m0,m0}^{1,1}$ are nearly superimposed on, respectively, $S_{m0,m0}^{1,2}$, $S_{m0,m0}^{2,2}$ for $m = 0, -1$ verifying the symmetry property of the element with respect to its median plan.
(2) The experiment underestimates the theoretical transmission coefficients for both modes near the cut-off frequency, for $1.5 < ka < 2$ and for ka close to 3. Also a slow increase with frequency from 0 at the cut-off frequency to 1 instead of the step function predicted by theory is evident for $m = -1$.
(3) The experimental reflection coefficients oscillate between 0 and 0.2 reaching higher values near cut-off frequencies in opposite to the theory which predicts no reflection.

(a) Experiment $|S_{00,00}^{12}|$: - - - - ; Experiment $|S_{00,00}^{21}|$: ———

Theory : — — —

(b) Experiment $|S_{00,00}^{22}|$: - - - - ; Experiment $|S_{00,00}^{11}|$: ———
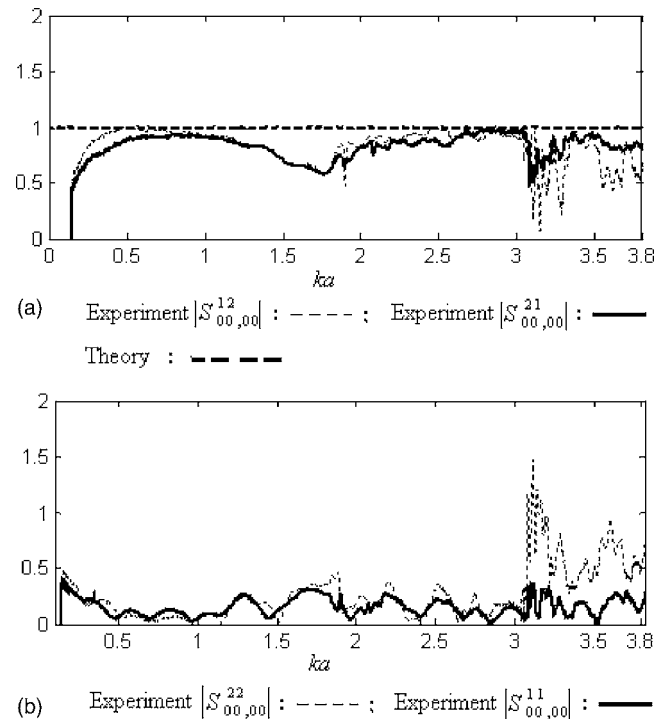
FIG. 9. Comparison with theory for the straight duct configuration of: (a) $|S_{00,00}^{2,1}|$ and $|S_{00,00}^{1,2}|$ the experimental transmission coefficients vs ka of the plane wave mode traveling, respectively, from the left to the right side and from the right to the left side; (b) $|S_{00,00}^{1,1}|$ and $|S_{00,00}^{2,2}|$ the experimental reflection coefficients vs ka of plane wave mode coming in the straight duct from the left side and from the right side, respectively.

(a) Experiment $|S^{21}_{-10,-10}|$: ——   Experiment $|S^{12}_{-10,-10}|$: ----

Theory : - - - -

(b)   Experiment $|S^{11}_{-10,-10}|$: —— ; Experiment $|S^{22}_{-10,-10}|$:----
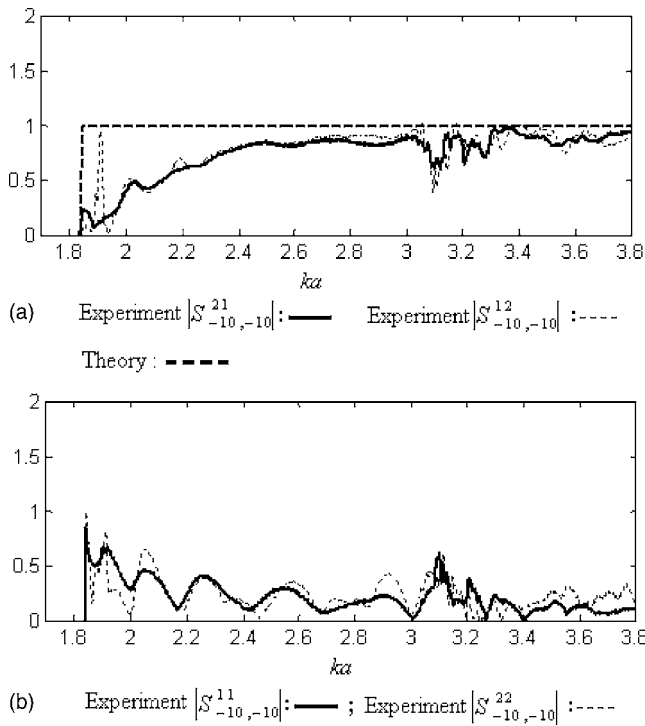
FIG. 10. For the straight duct configuration: (a) Comparison with theory of $|S^{2,1}_{-10,-10}|$ and $|S^{1,2}_{-10,-10}|$ the transmission coefficients vs ka of mode $(-1,0)$ traveling, respectively, from the left to the right side and from the right to the left side; (b) $|S^{1,1}_{-10,-10}|$ and $|S^{2,2}_{-10,-10}|$ the experimental reflection coefficients vs ka of mode $(-1,0)$ coming in the straight duct from the left side and from the right side, respectively.

The effect on the coefficients $S^{2,1}_{-10,-10}$ and $S^{1,1}_{-10,-10}$ of a 5° discrepancy on temperature and $(1°, 5\%)$ error on, respectively, the phase and the amplitude of modal pressure coefficients in two of the cross-section areas was estimated with the simulation procedure. In Fig. 11 the comparison between simulated and experimental results points out that these coefficients are very sensitive to errors but only in the frequency domain very close to the cut-off mode frequency. In Appendix D, this influence is shown to come from the sensitivity of the separation procedure in the input data such as temperature and total modal pressure explaining only part of the difference between experiment and theory. Indeed the periodic oscillation on the reflection coefficient and the slow variation of the transmission coefficient noticed before on mode $m=-1$ are not due to errors but caused by a physical phenomenon which is probably, as justified in the following, produced by a vibration of the duct wall. Indeed, the step variation of the theoretical transmission coefficient at the cut-off frequency and the nonreflection assume that the normal velocity on the wall is zero. But it is well known that this condition is not realistic and a coupling between acoustic duct cavity and duct wall vibration always occurs and can become stronger particularly when the frequency approaches by upper values the mode cut-off frequency.

Typical experimental results of the conversion coefficients of mode $m=-1$ into $m=0$ in transmission and reflection for waves coming in the discontinuity from the left and the right sides, respectively, are plotted in Fig. 12. All coefficients are lower than 0.1 except near the cut-off frequencies of modes. This result verifies the property of axisymmetry of the element which leads theoretically to no conversion between angular modes. Moreover, the couples of curves

(a) $|S^{2,1}_{-10,-10}|$: Experiment —— ; Simulation ------

(b) $|S^{1,1}_{-10,-10}|$: Experiment —— ; Simulation ------
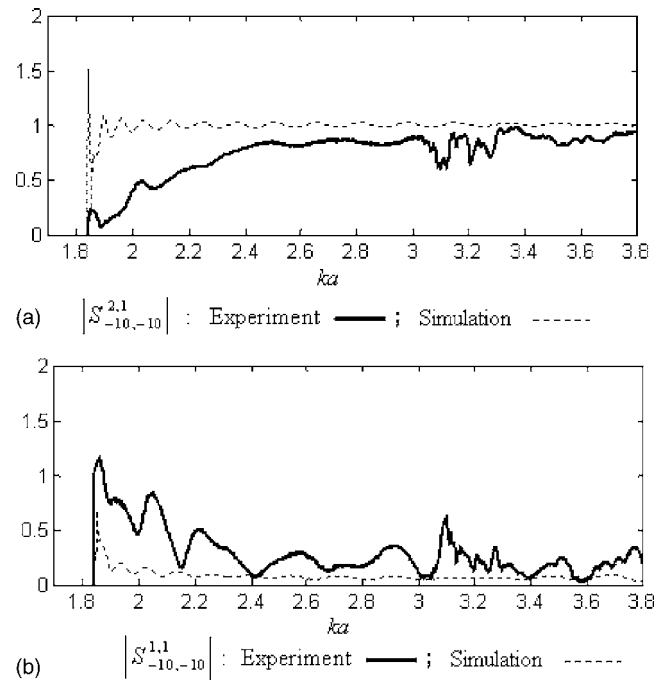
FIG. 11. For the straight duct configuration comparison between experiment and simulation with a 5° error on temperature and $(1°, 5\%)$ errors on, respectively, the phase and the amplitude of the modal pressure coefficient: (a) $|S^{2,1}_{-10,-10}|$ the transmission coefficient vs ka and (b) $|S^{1,1}_{-10,-10}|$ the reflection coefficient vs ka.

$[S^{2,1}_{00,-10}; S^{1,2}_{00,-10}]$ and $[S^{1,1}_{00,-10}; S^{2,2}_{00,-10}]$ are nearly superimposed verifying the assumption of symmetry of the element with respect to its median plan.

2. The Chamber

The modulus of the transmission coefficients $[S^{1,2}_{mn,mn}; S^{2,1}_{mn,mn}]$ and reflection coefficients $[S^{2,2}_{mn,mn}; S^{1,1}_{mn,mn}]$

(a)   Experiment $|S^{21}_{00,-10}|$: —— ; Experiment $|S^{12}_{00,-10}|$:----

(b)   Experiment $|S^{11}_{00,-10}|$: —— ; Experiment $|S^{22}_{00,-10}|$:----

FIG. 12. For the straight duct configuration the experimental conversion coefficients of mode $(-1,0)$ into $(0,0)$ vs ka: (a) $|S^{2,1}_{00,-10}|$ and $|S^{1,2}_{00,-10}|$ in transmission; (b) $|S^{1,1}_{00,-10}|$ and $|S^{2,2}_{00,-10}|$ in reflection.
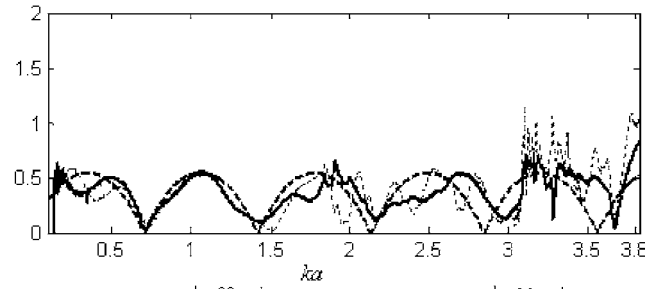
FIG. 13. Comparison for the reactive duct configuration between theory and experiment of: (a) $|S^{2,1}_{00,00}|$ and $|S^{1,2}_{00,00}|$ the transmission coefficients vs ka of mode $(0,0)$ traveling, respectively, from the left to the right side and from the right to the left side; (b) $|S^{1,1}_{00,00}|$ and $|S^{2,2}_{00,00}|$ the reflection coefficients of mode $(0,0)$ coming in the element, respectively, from the left side and from the right side.

FIG. 14. Comparison for the reactive duct configuration between theory and experiment of: (a) $|S^{2,1}_{+10,+10}|$ and $|S^{1,2}_{+10,+10}|$ the transmission coefficients vs ka of mode $(+1,0)$ traveling, respectively, from the left to the right side and from the right to the left side; (b) $|S^{1,1}_{+10,+10}|$ and $|S^{2,2}_{+10,+10}|$ the reflection coefficients vs ka of mode $(+1,0)$ coming in the element from the left side and from the right side, respectively.

through the duct element which contains the chamber for waves traveling from the left and the right sides are plotted for $m=0$, 1, and 2 and compared to the analytical results in Figs. 13, 14, and 16, respectively.

For the plane wave $m=0$ (Fig. 13) the agreement between experiment and theory is good but some differences have to be noticed:

(1) For $1.5 < ka < 2$ and around $ka=3$ the experimental transmission coefficient is smaller than the theoretical one as already observed before with the straight element and justified by a vibroacoustic coupling.

(2) The symmetry of the discontinuity with respect to its median plan is verified for $ka < 2$. The small diameter duct is filtering the pressure distributed on mode $m=1$ for $1.8 < ka < 2.51$ and $m=2$ for $3.05 < ka < 4.16$ (Table II) leading to a worse signal to noise ratio downstream of the chamber and then noisier coefficients $S^{1,2}_{00,00}$ and $S^{2,2}_{00,00}$ for waves traveling from the right to the left.

(3) Just below $ka=3.8$ the cut-off frequency of the mode $(0,1)$ the shift between experimental and theoretical minima and maxima is produced because the conversion of mode $(0,1)$ into $(0,0)$ is not taken into account by the analytical model.

The incident pressure $\boldsymbol{P}^{I+}_{10}(z)$ on mode $m=1$ on the source side is filtered by the chamber for $1.84 < ka < 2.51$ (Table II). Then in the analysis to follow the results associated to the waves traveling from the left and right sides of the chamber are discussed separately. Indeed Fig. 14 points

out the following:

(1) $S^{2,1}_{10,10}$ and $S^{1,1}_{10,10}$, the transmission and reflection coefficients for waves traveling from the left to the right, agree well with the theory.

(2) For ka up to 2.51 as $\boldsymbol{P}^{I+}_{10}(z)$ is filtered by the chamber, on the open side the pressure level is very low. The measurements of both coefficients $S^{1,2}_{10,10}$ and $S^{2,2}_{10,10}$ are based on the incoming wave traveling from the right side of the chamber. This pressure wave results from the reflection by the duct inlet of the wave which was already transmitted through the chamber. The pressure coming out on the left side of the chamber which is used to measure $S^{1,2}_{10,10}$ has then been filtered two times. This explains why in this frequency domain the signal to noise ratio becomes so low that such an important discrepancy occurs between experiment and theory and also why this error is more important on the transmission coefficient than on the reflection one. This effect is well reproduced (Fig. 15) by the simulation of the experiment where it was assumed a 5° discrepancy on temperature and 1°, 5% on,

TABLE II. Cut-off ka frequencies of duct with radius a and b.

(m,n)	ka ($a=7.48$ cm)	ka ($b=5.5$ cm)
$(0,0)$	0	0
$(\pm1,0)$	1.8412	2.51
$(\pm2,0)$	3.0542	4.16
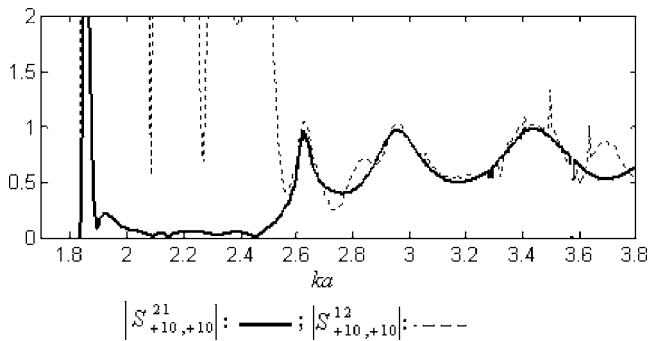
FIG. 15. For the reactive duct configuration simulation of $|S^{2,1}_{+10,+10}|$ and $|S^{1,2}_{+10,+10}|$ the transmission coefficients vs ka of mode $(+1,0)$ with a 5° error on temperature and $(1°,5\%)$ error on, respectively, the phase and the amplitude of the modal pressure coefficients.

respectively, the phase and the amplitude of modal pressure coefficients in the two cross sections. This result emphasizes also that the measurement of this coefficient in this propagation condition is very sensitive to errors on the experimental parameters.

(3) For $ka > 2.51$ the symmetry of the discontinuity with respect to its median plan is verified even for the plane wave case because the coefficients which are associated to the waves traveling from the right to the left are noisier than those traveling from the left to the right.

The incident pressure distributed on mode $(2,0)$ is cut-off by the chamber for all the ka spectrum studied (Table II). The interpretation of the experimental results (Fig. 16) developed for the mode $(1,0)$ when it is cut-off has to be extended for the $(2,0)$ mode to all ka studied. Indeed the trans-
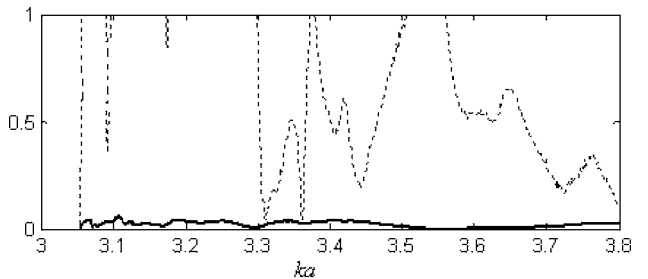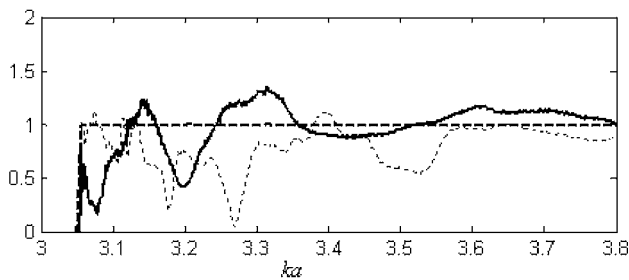
FIG. 16. Comparison for the reactive duct configuration with theory of: (a) $|S^{2,1}_{+20,+20}|$ and $|S^{1,2}_{+20,+20}|$ the experimental transmission coefficients vs ka of mode $(+2,0)$, respectively, from the left to the right side and from the right to the left side; (b) $|S^{1,1}_{+20,+20}|$ and $|S^{2,2}_{+10,+10}|$ the experimental reflection coefficients vs ka of mode $(+2,0)$ coming in the element from the left side and from the right side, respectively.
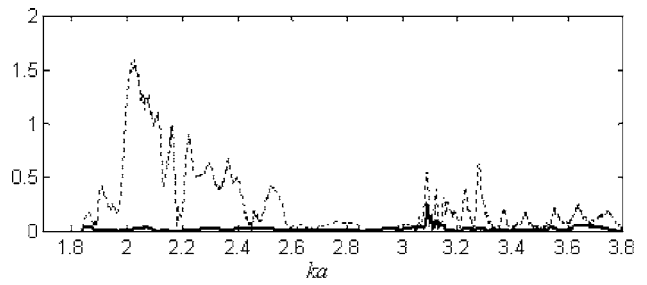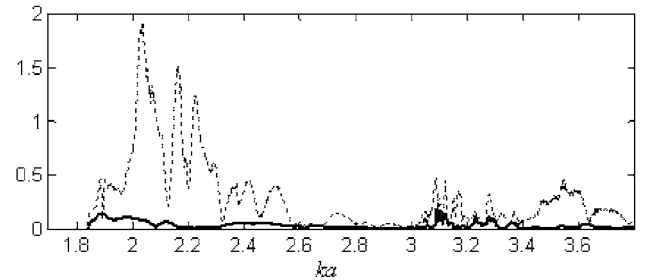
FIG. 17. For the reactive duct configuration, experimental conversion coefficients vs ka for mode $(-1,0)$ into $(0,0)$: (a) $|S^{2,1}_{00,-10}|$ and $|S^{1,2}_{00,-10}|$ transmission, respectively, from the left to the right side and from the right to the left side; (b) $|S^{1,1}_{00,-10}|$ and $|S^{2,2}_{00,-10}|$ reflection coming in the element from the left side and from the right side, respectively.

mission coefficient $S^{2,1}_{20,20}$ and the reflection one $S^{1,1}_{20,20}$ related to the wave traveling from the left to the right agree with theory. But the measurement of $S^{1,2}_{20,20}$ transmission coefficient resulting from the wave traveling from the right to the left is completely wrong while the reflection coefficient $S^{2,2}_{20,20}$ is better determined allowing one to verify the symmetry of the discontinuity with respect to the chamber median plan.

In Fig. 17 are plotted conversion coefficients in transmission and reflection of the mode $(-1,0)$ into the plane wave. Again the coefficients agree with theory, which predicts a zero value because of the axisymmetry of the geometry but only for the wave traveling from the left to the right. The computation by the simulation program of the conversion coefficients in reflection with errors on temperature and modal coefficients reproduces the experimental curves as shown in Fig. 18, verifying the high sensitivity of this measurement to uncertainties.

In Fig. 19, the results of the measurement of $|S^{2,1}_{10,10}|$, the modulus of the transmission coefficient versus ka already presented in Fig. 14, are compared with the results deduced when the modal basis used for the calculation of $[S]$ includes all five modes in the complete frequency domain meaning that the influence of the evanescent mode is taken into account. As both curves are superimposed, we can deduce that in our test configuration the nonpropagating modes can be neglected.

V. CONCLUSIONS

An experimental procedure was achieved to measure the scattering matrix of a duct discontinuity for higher order mode propagation conditions. Straight and reactive type silencer duct configurations were tested and the experimental
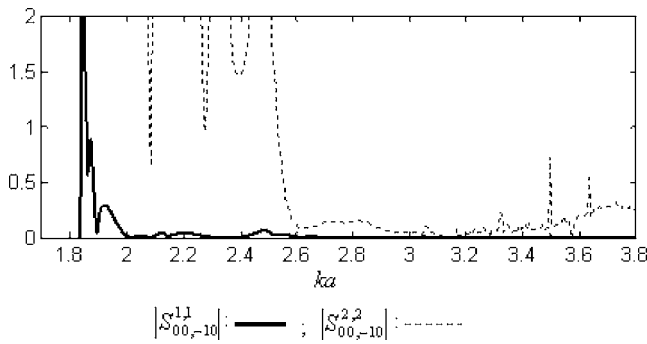
FIG. 18. For the reactive duct configuration, simulation with a 5° error on temperature and $(1°, 5\%)$ errors on, respectively, the phase and the amplitude of the modal pressure coefficients of $|S_{00,-10}^{1,1}|$ and $|S_{00,-10}^{2,2}|$ the conversion coefficients vs ka in reflection for mode $(-1,0)$ into $(0,0)$ respectively coming in the element from the left side and from the right side.

results compared with theory. The analysis of the results performed mode per mode on the straight configuration showed that the overdetermination procedure with six loads improves the results. An analytical procedure of the experiment was developed and used to point out the sensitivity of the experimental parameters errors on $[S]$ measurements. Indeed, it was shown to be important but localized very near the mode cut-off frequency. The other differences between theoretical and experimental variations of the transmission and reflection coefficients are explained by the assumption of the non-realistic theoretical hard duct wall condition. The symmetry properties of the straight duct geometry with respect to the duct axis and its medium plan were verified experimentally.

The experiment with the reactive chamber has pointed out the filtering effect of a reduction of the diameter of the duct on acoustic modal propagation. Experimental and theoretical results agree in the frequency domain where the modes are cut-on in all test configurations. When the modes are cut-off by the smaller diameter duct the experimental results associated to the incident wave coming from the source side also agree with theory. But the method is unable to measure $[S]$, especially the transmission coefficient associated to the wave traveling from the open end side in this frequency domain. The simulation also shows that, in these conditions, the method is very sensitive to errors leading clearly to wrong results. When the discontinuity is not symmetric with respect to its middle plan, the two source technique or a permutation of the chamber has to be applied to improve the results.
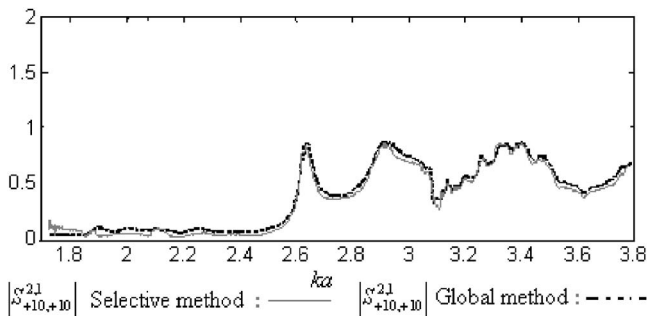
FIG. 19. For the reactive duct configuration experimental $|S_{+10,+10}^{2,1}|$ the transmission coefficient vs ka of mode $(+1,0)$ traveling from the left to the right side deduced from selective and global methods.

APPENDIX A: CALCULATION OF THE SCATTERING MATRIX FROM A GIVEN TRANSFER MATRIX

The transfer matrix $[T]_{N\times N}$ of a duct element like this shown in Fig. 1 is defined by

$$\begin{Bmatrix} P^{II} \\ V_z^{II} \end{Bmatrix}_{2N} = [T]_{2N\times 2N} \times \begin{Bmatrix} P^{I} \\ V_z^{I} \end{Bmatrix}_{2N}. \qquad (A1)$$

P and V_z are, respectively, the total pressure and the axial acoustic particle velocity vectors expressed versus their modal description by

$$\{P^{I}, V_z^{I}\}_{2N} = \{P_{mn}^{I}(z_L), V_{z,mn}^{I}(z_L)\}_N^t, \qquad (A2)$$

$$\{P^{II}, V_z^{II}\}_{2N} = \{P_{mn}^{II}(z_R), V_{z,mn}^{II}(z_R)\}_N^t.$$

It is well known that the axial acoustic particle velocity is deduced from the pressure by[1]

$$V_{z,mn}(z) = Y_{mn}^a (P_{mn}^+(z) - P_{mn}^-(z)) \qquad (A3)$$

with $Y_{mn}^a = (k_{mn}^a / \rho_0 c_0 k)$.

Developing Eq. (A1) and replacing these total acoustic quantities versus the incident $(+)$ and reflected $(-)$ pressure and axial particle velocities two relationships are deduced:

$$\{P_{mn}^{II+}(z_R)\}_N + \{P_{mn}^{II-}(z_R)\}_N = [X^+]_{N\times N} \cdot \{P_{pq}^{I+}(z_L)\}_N$$
$$+ [X^-]_{N\times N} \cdot \{P_{pq}^{I-}(z_L)\}_N, \qquad (A4)$$

$$\{P_{mn}^{II+}(z_R)\}_N - \{P_{mn}^{II-}(z_R)\}_N = [W^+]_{N\times N} \cdot \{P_{pq}^{I+}(z_L)\}_N$$
$$+ [W^-]_{N\times N} \cdot \{P_{pq}^{I-}(z_L)\}_N, \qquad (A5)$$

where

$$[X^\pm]_{N\times N} = ([T^{11}]_{N\times N} \pm [\text{diag}(Y_{mn}^a)]_{N\times N} \cdot [T^{12}]_{N\times N}), \qquad (A6)$$

$$[W^\pm]_{N\times N} = [\text{diag}(Y_{mn}^a)]_{N\times N}^{-1} \cdot ([T^{21}]_{N\times N}$$
$$\pm [\text{diag}(Y_{mn}^a)]_{N\times N} \cdot [T^{22}]_{N\times N}). \qquad (A7)$$

As the scattering matrix is given by Eq. (1) and using Eqs. (A5) and (A6) the expressions of the four elementary matrices of $[S]$ vs $[X^\pm]_{N\times N}$ and $[W^\pm]_{N\times N}$ are deduced:

$$[S^{1,1}]_{N\times N} = [[[X^-]_{N\times N} - [W^-]_{N\times N}]^{-1} \cdot ([W^+]_{N\times N}$$
$$- [X^+]_{N\times N})]_{N\times N}, \qquad (A8)$$

$$[S^{1,2}]_{N\times N} = 2[[X^-]_{N\times N} - [W^-]_{N\times N}]_{N\times N}^{-1}, \qquad (A9)$$

$$[S^{2,2}]_{N\times N} = [-[[X^-]_{N\times N}^{-1} - [W^-]_{N\times N}^{-1}]^{-1} \cdot ([X^-]_{N\times N}^{-1}$$
$$+ [W^-]_{N\times N}^{-1})]_{N\times N}, \qquad (A10)$$

$$[S^{2,1}]_{N\times N} = [[[X^-]_{N\times N}^{-1}$$
$$- [W^-]_{N\times N}^{-1}]^{-1} \cdot ([X^-]_{N\times N}^{-1} \cdot [X^+]_{N\times N}$$
$$- [W^-]_{N\times N}^{-1} \cdot [W^+]_{N\times N})]_{N\times N}. \qquad (A11)$$
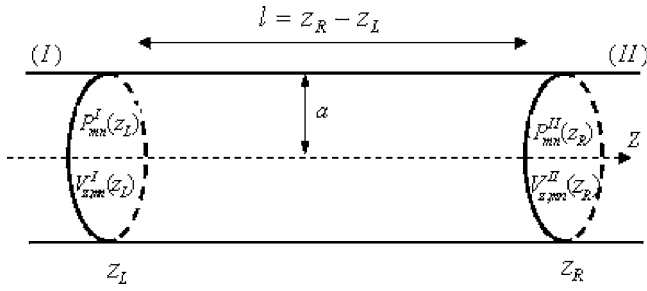
FIG. 20. Schema of the straight duct configuration.

FIG. 21. Schema of the reactive chamber configuration.

APPENDIX B: CALCULATION OF THE TRANSFER MATRICES OF THE STRAIGHT DUCT AND OF THE REACTIVE CHAMBER CONFIGURATIONS

The scattering matrices of both configurations are deduced from their transfer matrices using the relationships given in Appendix A. Then in the following paragraphs only the transfer matrices of both configurations are calculated.

The straight duct (Fig. 20). If N modes are cut-on in a straight duct with radius a located between z_L and z_R, its transfer matrix is defined[1] also by Eqs. (A1) and (A2). If $V_{z,mn}(z)$ is given by Eq. (A3), the transfer matrix of a straight element of radius a and length l is[1]

$$[T_{a,l}]_{2N\times 2N} = \begin{bmatrix} [\mathrm{diag}(\cos(k_{mn}^a l))]_{N\times N} & [\mathrm{diag}((i/Y_{mn}^a)\cdot\sin(k_{mn}^a l))]_{N\times N} \\ [\mathrm{diag}(iY_{mn}^a\cdot\sin(k_{mn}^a l))]_{N\times N} & [\mathrm{diag}(\cos(k_{mn}^a l))]_{N\times N} \end{bmatrix}_{2N\times 2N}. \tag{B1}$$

Reactive muffler (Fig. 21). If N modes are cut-on in the main duct with radius a, the transfer matrix of the duct element located between z_L and z_R is deduced by

$$[T]_{2N\times 2N} = [T_{a,l_a}]_{2N\times 2N}[T^{\mathrm{Const}}]_{2N\times 2N}$$
$$\times [T_{b,l_b}]_{2N\times 2N}[T^{\mathrm{Exp}}]_{2N\times 2N}[T_{a,l_a}]_{2N\times 2N}. \tag{B2}$$

$[T_{a,l_a}]_{2N\times 2N}$ is the transfer matrix of straight duct with radius a and length l_a. $[T_{b,l_b}]_{2N\times 2N}$ is the transfer matrix of straight duct with radius b and length l_b.

The transfer matrix of an expansion $[T^{\mathrm{Exp}}]_{2N\times 2N}$ is deduced by inversing the transfer matrix of the constriction matrix given by

$$[T^{\mathrm{Const}}]_{2N\times 2N} = \begin{bmatrix} [F]_{N\times N} & [0]_{N\times N} \\ [0]_{N\times N} & [G]_{N\times N}^{-1} \end{bmatrix}_{2N\times 2N}. \tag{B3}$$

$[F]_{N\times N}$ and $[G]_{N\times N}$ are deduced from the continuity boundary conditions on pressure and axial particle velocity in the duct crosssection S where the area modification occurs:[8]

$$F_{pq,mn} = \frac{1}{N_{pq}^b}\int\int_{S_b}\Psi_{mn}^a\Psi_{pq}^{b*}rdrd\theta, \tag{B4}$$

$$G_{mn,pq} = \frac{1}{N_{mn}^a}\int\int_{S_a}\Psi_{pq}^b\Psi_{mn}^{a*}rdrd\theta.$$

$N_{mn}^{b,a} = S_{b,a}J_m^2(\chi_{mn})[1-(m^2/\chi_{mn}^2)]$ is the normalization factor. $S_a = \pi(a)^2$ and $S_b = \pi(b)^2$.

APPENDIX C: COMPUTATION OF THE PRESSURE IN THE TWO PAIRS OF DUCT CROSS SECTIONS DURING THE SIMULATION PROCEDURE OF THE EXPERIMENT

The expression of the incident pressure $\{P_{mn}^{\mathrm{I}+}(z_L)\}_N$ distributed on N cut-on modes produced by a point source (Fig. 6) in the semi-infinite duct at z_L is given by[8]

$$\{P_{mn}^{\mathrm{I}+}(z_L)\}_N = \{\alpha_{mn}A_S e^{i(\varphi_S - m\theta_S + k_{mn}(z_L - z_S))}\}_N, \tag{C1}$$

where $\alpha_{mn} = 1/2ik_{mn}N_{mn}^a$, $A_S = 1$, and $\varphi_S = 0$ are, respectively, the phase and amplitude of the point source, θ_S and z_S are angular and axial positions of the point source.

The transmission matrix $[T_R]_{N\times N}$ of the duct element (Fig. 6) located between z_L and z_R is defined as a function of the incident and transmitted pressures distributed on the cut-on modes of the main duct by[9]

$$\{P_{mn}^{\mathrm{II}+}(z_R)\}_N = [T_R]_{N\times N}\cdot\{P_{mn}^{\mathrm{I}+}(z_L)\}_N. \tag{C2}$$

If $[R^{\mathrm{I}}(z_L)]_{N\times N}$ and $[R^{\mathrm{II}}(z_R)]_{N\times N}$ are the reflection matrices in both cross-section areas located at z_L and z_R apart from the discontinuity (Fig. 7) the incident and reflected pressures in z_L and z_R distributed on cut-on modes of the main duct are related by

$$\{P_{mn}^{\mathrm{II}-}(z_R)\}_N = [R^{\mathrm{II}}(z_R)]_{N\times N}\cdot\{P_{mn}^{\mathrm{II}+}(z_R)\}_N \quad\text{and} \tag{C3}$$

$$\{P_{mn}^{\mathrm{I}-}(z_L)\}_N = [R^{\mathrm{I}}(z_L)]_{N\times N}\cdot\{P_{mn}^{\mathrm{I}+}(z_L)\}_N.$$

The reflection matrix in z_R is deduced from the reflection matrix $[R^{\mathrm{II}}(z_{\mathrm{end}})]_{N\times N}$ in the open cross-section area located at z_{end} computed from a theoretical model which assumes an infinite baffle condition.[21]
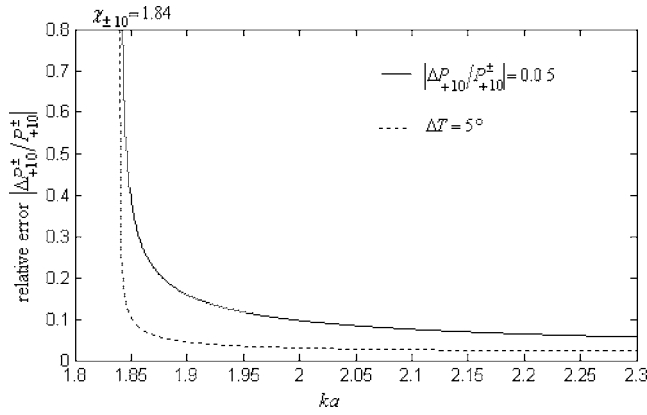
FIG. 22. The relative uncertainty vs ka on the determination of the incident $(1,0)$ modal pressure by the separation technique to $5°$ in T or 5% in modulus of total modal coefficient.

FIG. 23. The relative uncertainty vs ka on the determination of the incident plane wave pressure by the separation technique to $5°$ in T or 5% in modulus of total modal coefficient.

$$[R^{\text{II}}(z_R)]_{N\times N} = [\text{diag}(e^{-ik_{mn}E})]_{N\times N}^{-1} \cdot [R^{\text{II}}(z_{\text{end}})]_{N\times N}$$
$$\times [\text{diag}(e^{+ik_{mn}E})]_{N\times N}. \tag{C4}$$

$[T_R]_{N\times N}$ and $[R^{\text{I}}(z_L)]_{N\times N}$ are deduced from the scattering matrix for a given length E load with the relationships:

$$[T_R]_{N\times N}$$
$$= [[I]_{N\times N} - [S^{22}]_{N\times N} \cdot [R^{\text{II}}(z_R)]_{N\times N}]^{-1}_{N\times N} \cdot [S^{21}]_{N\times N}, \tag{C5}$$

$$[R^{\text{I}}(z_L)]_{N\times N}$$
$$= [[S^{11}]_{N\times N} + [S^{12}]_{N\times N} \cdot [R^{\text{II}}(z_R)]_{N\times N} \cdot [T_R]_{N\times N}]_{N\times N}. \tag{C6}$$

The total acoustic pressures in the four cross sections located at $z_L, z_{L+d}, z_R,$ and z_{R+d} are then deduced by adding the incident and reflected pressures:

$$\{P_{mn}^{\text{I,II}}(z_{L,R})\}_N = \{P_{mn}^{\text{I,II}+}(z_{L,R})\}_N + \{P_{mn}^{\text{I,II}-}(z_{L,R})\}_N, \tag{C7}$$

$$\{P_{mn}^{\text{I,II}}(z_{L,R}+d)\}_N = [\text{diag}(e^{+ik_{mn}d})]_{N\times N} \cdot \{P_{mn}^{\text{I,II}+}(z_{L,R})\}_N$$
$$+ [\text{diag}(e^{-ik_{mn}d})]_{N\times N} \cdot \{P_{mn}^{\text{I,II}-}(z_{L,R})\}_N. \tag{C8}$$

This calculation is performed for $q=1$, N source configurations $(z_S, \theta_S)^q$ and all loads (Table I).

APPENDIX D: ANALYTIC SOLUTION OF THE ERROR CALCULATION ISSUED FROM THE SEPARATION TECHNIQUE

The complex incident and reflected modal pressures P_{mn}^+ and P_{mn}^- are solutions of the following system[19] where $P_{mn}^{1,2}$ are the total modal pressures in two cross sections separated by a distance d (Fig. 2):

$$P_{mn}^+ = \frac{P_{mn}^1 e^{-ik_{mn}d} - P_{mn}^2}{e^{-ik_{mn}d} - e^{ik_{mn}d}},$$

$$P_{mn}^- = \frac{P_{mn}^2 - P_{mn}^1 e^{ik_{mn}d}}{e^{-ik_{mn}d} - e^{ik_{mn}d}}. \tag{D1}$$

If $\Delta P_{mn}^{1,2}$, Δk_{mn}, and Δd are, respectively, the uncertainties on the total modal pressures $P_{mn}^{1,2}$, axial wave number, and distance, the error on $P_{mn}^\pm$ can be deduced from

$$\Delta P_{mn}^\pm \simeq \xi_{mn}^\pm \Delta P_{mn} + \eta_{mn}^\pm(d\Delta k_{mn} + k_{mn}\Delta d), \tag{D2}$$

where

$$\xi_{mn}^\pm = \frac{\partial P_{mn}^\pm}{\partial P_{mn}^1} + \frac{\partial P_{mn}^\pm}{\partial P_{mn}^2} = \frac{\pm e^{\mp ik_{mn}d} \pm 1}{e^{-ik_{mn}d} - e^{ik_{mn}d}},$$

$$\eta_{mn}^\pm = \frac{\partial P_{mn}^\pm}{\partial(k_{mn}d)} = \pm 2 \cdot i\left(\frac{P_{mn}^1 - P_{mn}^2\cos(k_{mn}d)}{(e^{-ik_{mn}d} - e^{ik_{mn}d})^2}\right).$$

Assuming a reflecting condition where the conversion is neglected:

$$P_{mn}^- = P_{mn}^+ R_{mn,mn} \Rightarrow \begin{cases} P_{mn}^1 = P_{mn}^+(1 + R_{mn,mn}) \\ P_{mn}^2 = P_{mn}^+(e^{ik_{mn}d} + e^{-ik_{mn}d}R_{mn,mn}) \end{cases}$$

the coefficient $\eta_{mn}^\pm$ becomes

$$\eta_{mn}^\pm = \pm 2iP_{mn}^+\left(\frac{(1 - e^{ik_{mn}d}\cos(k_{mn}d)) + R_{mn,mn}(1 - e^{-ik_{mn}d}\cos(k_{mn}d))}{(e^{-ik_{mn}d} - e^{ik_{mn}d})^2}\right).$$

If the hard wall reflection condition $|R_{mn,mn}| = 1$ and $\Phi(R_{mn,mn}) = 0$ is assumed, $\eta_{mn}^\pm = \pm iP_{mn}^+$ and Eq. (D2) is leading to the relationship which gives the error on the incident and reflected modal pressure relative to the incident modal pressure:

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Sitel *et al.*: Acoustic scattering matrix measurement   2489

$$\Delta P_{mn}^{\pm} / P_{mn}^{\pm} \simeq \xi_{mn}^{\pm} \cdot (\Delta P_{mn} / P_{mn}^{\pm}) + i(d \cdot \Delta k_{mn} + k_{mn} \cdot \Delta d). \tag{D3}$$

The uncertainty Δk_{mn} on the axial wave number k_{mn} is deduced from the uncertainties on T the temperature and on f_{mn}^c cut-off frequency of mode (m,n) by

$$\Delta k_{mn} \simeq \frac{\partial k_{mn}}{\partial c_0} \frac{\partial c_0}{\partial T} \Delta T + \frac{\partial k_{mn}}{\partial f_{mn}^c} \Delta f_{mn}^c = \mu_{mn}' \Delta T + \mu_{mn} \Delta f_{mn}^c, \tag{D4}$$

where $c_0 = 20\sqrt{273+T} \simeq 330.45 + 0.607T$. As

$$f_{mn}^c = \frac{c_0}{2\pi} \frac{\chi_{mn}}{a}, \text{ it is easy to show that}$$

$$\Delta f_{mn}^c \simeq 0.607 \frac{\chi_{mn}}{2\pi a} \Delta T + \frac{c_0}{2\pi a} \Delta \chi_{mn} + \frac{\chi_{mn} c_0}{2\pi a^2} \Delta a, \tag{D5}$$

where $\mu_{mn} = (2\pi/c_0)^2 f_{mn}^c / k_{mn}$ and $\mu_{mn}' \simeq 0.607 k_{mn}/c_0$. $\Delta \chi_{mn}$ is an error issued from the measurement duct wall boundary condition and Δa the error on the measurement duct radius.

From the relationships (D3) and (D4) the relative uncertainty $\Delta P_{mn}^+ / P_{mn}^+$ to errors in T, χ_{mn}, a can then be deduced. The contributions to the relative error $\Delta P_{mn}^+ / P_{mn}^+$ vs ka of a $5°$ error in temperature and 5% error in the modulus of the total modal pressure are plotted in Fig. 22 for a mode $(1,0)$ and Fig. 23 for the plane wave. These curves point out that for the plane case the error on the pressure measurement is more important than this on the temperature as shown in a previous study[20] while for a higher order mode both contributions are similar in amplitude and become very sensitive when the frequency tends to be near to the cut-off mode frequency.

[1] M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley-Interscience, New York, 1987).

[2] P. O. A. L. Davies, "Practical flow duct acoustics," J. Sound Vib. **124**, 91–115 (1988).

[3] R. Glav and M. Aböm, "A general formalism for analyzing acoustic 2-port network," J. Sound Vib. **202**(5), 739–747 (1997).

[4] W. Evermans, "A systematic procedure for the analysis of multiply branched acoustic transmission lines," Trans. ASME, J. Vib., Acoust., Stress, Reliab. Des. **109**, 168–177 (1987).

[5] Y. Auregan and R. Starobinski, "Determination of acoustical energy dissipation/production potentiality from the acoustical transfer functions of a multiport," Acust. Acta Acust. **85**, 788–792 (1999).

[6] Z. Tao and A. F. Seybert, "A review of current techniques for measuring mufflers transmission loss," Soc. Auto. Eng. 03NVC-38, 2001.

[7] M. Abom, "Measurement of the scattering matrix of acoustical two-ports," Mech. Syst. Signal Process. **5**(2), 89–104 (1991).

[8] M. Akoum and J. M. Ville, "Measurement of reflection matrix of a discontinuity in a duct," J. Acoust. Soc. Am. **103**(5), 2463–2468 (1998).

[9] A. Sitel, J. M. Ville, and F. Foucart, "An experimental facility for measurement of acoustic transmission matrix and acoustic power dissipation of duct discontinuity in higher order modes propagation conditions," Acust. Acta Acust. **89**, 586–594 (2003).

[10] R. T. Muehleisen and D. C. Swanson, "Modal coupling in acoustic waveguides: Planar discontinuities," Appl. Acoust. **63**, 1375–1392 (2002).

[11] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[12] E. R. Rademaker *et al.*, Publishable Synthesis Report No. DUCAT-NL-01, 2001.

[13] J. M. Auger and J. M. Ville, "Measurement of liner impedance based on the determination of duct eigenvalues by Fourier-Lommel's transform," J. Acoust. Soc. Am. **88**, 19–22 (1990).

[14] H. Boden and M. Abom, "Influence of errors on the two-microphone method for measuring acoustic properties in ducts," J. Acoust. Soc. Am. **79**(2), 541–549 (1985).

[15] J. Lavrenjev and M. Äbom, "Characterisation of fluid machines as acoustic multi-port sources," J. Sound Vib. **197**(1), 1–16 (1996).

[16] H. Bodèn, "The multiple load method for measurement of the source characteristics of time-variant sources," J. Sound Vib. **148**(3), 437-453 (1991).

[17] G. H. Golub and C. F. Loan, *Matrix Computations* (North Oxford Academic, Oxford, 1983).

[18] J. M. Auger and J. M. Ville, "Flow effects on measurement of the modal decomposition of acoustic field in a hard cylindrical duct," Proceedings of the Aero and Hydro Acoustics IUTAM Symposium, Lyon, France, 1985, pp. 437–443.

[19] M. Abom, "Modal decomposition in ducts based on transfer function measurements between microphone pairs," J. Sound Vib. **135**(1), 95–114 (1989).

[20] R. T. Muehleisen and C. W. Beamer IV, "Comparison of errors in the three and four microphone method used in the measurement of acoustic properties of porous materials," ARLO **3**(4), 112–117 (2002).

[21] W. E. Zorumski, "Generalized radiation impedance and reflection coefficients of circular and annular ducts," J. Acoust. Soc. Am. **54**, 1667–1673 (1973).

Measurement and prediction of nonlinearity in outdoor propagation of periodic signals

Kent L. Gee[b)] and Victor W. Sparrow
Graduate Program in Acoustics, 202 Applied Science Bldg., The Pennsylvania State University, University Park, Pennsylvania 16802

Michael M. James,[c)] J. Micah Downing,[c)] and Christopher M. Hobbs
Wyle Laboratories, 2001 Jefferson Davis Hwy. Ste. 701, Arlington, Virginia 22202

Far field propagation measurements of high-amplitude periodic signals generated by the U. S. Army Research Laboratory's Mobile Acoustic Source (MOAS) have been made. The MOAS is a large horn-coupled electropneumatic loudspeaker capable of producing sound at a few hundred hertz with a maximum overall sound pressure level of 155 dB re 20 μPa at 1 m. The possible influence of nonlinear effects have been investigated because the measurements exhibit greater sound pressure levels at high harmonics than are predicted by a linear propagation model. Between 100 and 375 m, nonlinearly predicted spectra obtained via a generalized Burgers equation-based model are consistently closer to measured spectra than are linear predictions, according to calculations of mean absolute error. These comparisons strengthen the assertion that nonlinearity is, in fact, the primary cause of disagreement between the measured and linearly predicted spectra at high frequencies. Comparisons between the nonlinear model and measurements, however, yield increased errors for greater propagation distances ($\sim$1 km) and for measurements made later in the afternoon. For these cases, the nonlinear model calculations generally predict greater sound pressure levels at high frequencies than are actually present in the MOAS measurements. Despite the increased errors for these latter comparisons, the nonlinear model still typically performs better than the linear model. This provides additional confirmation of the presence of nonlinearity in the propagation, but may also point to the need to account for atmospheric variability in the numerical model to provide improved predictions. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2345934]

I. INTRODUCTION

In the past, the role of nonlinearity in the propagation of high-amplitude sound has been given considerable attention in a variety of contexts. Relevant studies have utilized analytical, numerical, and experimental methods to study various aspects of the problem. However, a review of the literature reveals only a limited number of experiments dedicated to the measurement of finite-amplitude effects in outdoor continuous-wave sound propagation. Theobald[1] studied the vertical propagation of periodic waveforms over a maximum range of 76 m. Webster and Blackstock[2] subsequently performed a similar study over a comparable range with band-limited noise waveforms. Both of these studies showed clear evidence of nonlinear propagation in that the measured high-frequency sound pressure levels were significantly greater than those predicted with linear theory.

Our main purpose in this article is to describe the results of recent field experiments with a high-amplitude acoustic source that also show evidence of nonlinear effects. The source for these measurements was the U. S. Army Research Laboratory's Mobile Acoustic Source[3] (MOAS), which is a large electropneumatic loudspeaker. These experiments with the MOAS have been carried out as part of a multifaceted effort to study nonlinearity in the propagation of noise from high-performance jet aircraft[4] over moderate ranges. Because of the potential impact that takeoffs and low-altitude training runs may have on nearby communities, nonlinearity in ground-to-ground propagation is of considerable interest. Consequently, these controlled-source measurements were conducted over several hundred meters at near-grazing incidence, for which ground effects can play a considerable role. Also, as is typical with outdoor propagation experiments, wind, turbulence, and temperature profiles affect the recorded waveforms.

Because multiple phenomena influence the propagation, an important aspect of the measurement analysis is a comparison against the results of a numerical model[5] that is based on the generalized Burgers equation (GBE). The GBE is a widely established nonlinear model equation that can correctly predict the effects of second-order nonlinearity for lossy parabolic propagation of acoustic pressure waveforms. However, because the GBE does not incorporate all the phenomena that affect outdoor measurements, a comparison of

FIG. 1. (Color online) Photograph of the U.S. Army Research Laboratory's Mobile Acoustic Source (MOAS).

FIG. 2. (Color online) Microphone ranges and heights for the MOAS propagation measurements.

the numerical results against the measurement may give some indication of the relative importance of nonlinear effects at a given frequency for a particular test. Theobald[1] also made numerical comparisons against his periodic-waveform measurements using a modified form of an algorithm developed by Pestorius and Blackstock.[6] Agreement between measured versus numerically calculated levels for the first three harmonics was good over short distances, but significantly worsened for the second and third harmonics by the maximum measurement distance of 76 m. Computed results were obtained using only one cycle of the input waveform that was scaled to match the time-averaged level of the fundamental frequency.

In the remainder of this paper, the MOAS and the measurement setup are first described. Measurement results for a number of cases are then presented and discussed. Finally, after a brief discussion of the numerical model, comparisons between measured and both nonlinearly and linearly predicted spectra are examined and analyzed.

II. MEASUREMENT SUMMARY

A. The mobile acoustic source

The U. S. Army Research Laboratory has a large horn-coupled electropneumatic loudspeaker known as the Mobile Acoustic Source (MOAS). The horn is a 10-Hz exponential horn that is 17.1 m long and has a mouth diameter of 2.3 m. Acoustic pressure signals are generated by a valve that modulates the flow of compressed air. The MOAS, which has a nominal frequency response range of 10–500 Hz, was primarily designed to simulate the acoustical signature of a tank. Measurements of the MOAS conducted by the National Center for Physical Acoustics at the University of Mississippi[7] indicate maximum output levels of 155 dB *re* 20 μPa at 1 m. The MOAS is shown mounted on its flatbed trailer in Fig. 1, where the horn's axis is located 3 m above the ground.

B. Measurement array and environment

The propagation measurements were conducted on 12 February 2004 at the Blossom Point field test facility in Blossom Point, MD between 13:00 and 14:45 Eastern Standard Time (EST). Bruel and Kjaer 12.7-mm free-field microphones (Type 4190) were located along the loudspeaker centerline according to the layout in Fig. 2. Microphone poles were placed at 10, 100, 250, and 375 m, and an existing tower was used to collect data at 1092 m. The measurement elevation angles were 0°, 0.7°, and 1.4°, relative to the 10-
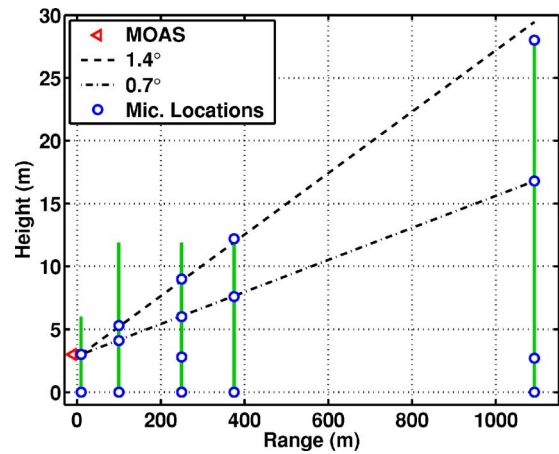
m on-axis microphone. The 1.4° measurement angle is approximate because the tower height was insufficient to place the 1092-m microphone along the same angle as the rest of the microphones, which were located nominally at 1.45°. Microphones were also placed on the ground at each pole location and at the tower. All microphones were pointed at the source (normal incidence), which yields a nominally flat (±1 dB) amplitude response for the Type 4190 microphones out to 20 kHz. With the exception of the tower data at 1092 m, microphone data acquisition was carried out using National Instruments 24-bit PXI-4472 cards sampling at 96 kHz. At the tower, waveform data were acquired with a National Instruments 16-bit NI-DAQ 6036-E card sampling at 200 kHz and were then multiplexed over four channels, yielding an effective sampling frequency of 50 kHz per channel.

The ground along the propagation path was fairly flat, nominally soft, and covered by long grass and other vegetation. Because of hardware failure, meteorological information at the site was not recorded during the measurements. However, data from the Naval Surface Warfare Center Dahlgren laboratory located approximately 10 km (6 mi) from the measurement site provided a reasonable estimate of average weather conditions based on a comparison between Blossom Point and Dahlgren data made the day before.

According to the Dahlgren data, between 13:00 and 14:45 EST, the ambient pressure was approximately constant at 1.0 atm, the temperature range was 4 °–5 °C, and the relative humidity varied between 70% and 77%. Wind speeds ranged between 1 and 6 m/s with a variable direction. Because of the uncertainty in the ambient conditions, the values for atmospheric absorption and dispersion calculations in the propagation model were assumed to be constant at 1 atm for ambient pressure, 4 °C for temperature, and 73% for relative humidity.

C. Measured waveforms

For the propagation measurements, a variety of periodic waveforms with fundamental frequencies ranging from 50–400 Hz were used as signal inputs. Sine, triangle, and square waves were all used, but the MOAS frequency re-
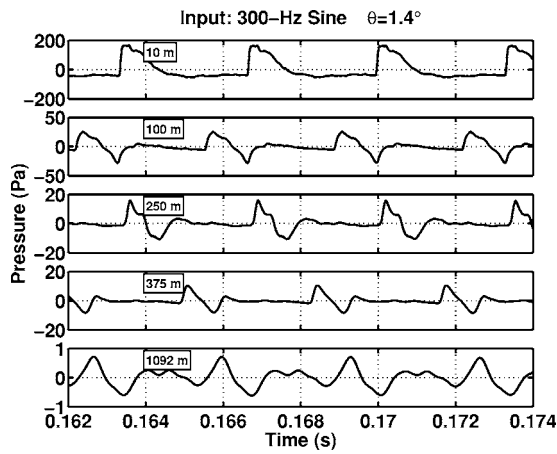
FIG. 3. Measured MOAS waveforms along 1.4° for a 300-Hz input sine wave. Distance labels correspond to the microphone ranges shown in Fig. 2.

FIG. 4. (Color online) Measured MOAS harmonic levels along 1.4° for a 300-Hz input sine wave. In this as well as subsequent figures, OASPL signifies overall sound pressure level and is referenced to 20 μPa.

sponse and inherent distortion were such that its ability to acoustically transmit the supplied electrical signals was rather poor. Consequently, the type of input waveform was later found to be largely irrelevant. In addition to the periodic signals described, pink noise was also used. However, the limited frequency response of the MOAS resulted in a natural bandpass filtering of the noise signal and, consequently, significantly lower amplitudes for the pink noise tests.

As an example of a typical measurement, recorded waveforms along 1.4° for a 300-Hz sine wave input are shown in Fig. 3. Although a sinusoidal electrical signal was supplied, the resultant waveform at 10 m is quite nonsinusoidal and significantly skewed, but it is periodic with a fundamental frequency of 300 Hz. The waveform skewness at close range could be related to the compressed-air nature of the source that results in a net volume increase over a cycle. Similarly skewed, nonsinusoidal waveforms recorded at close range for a 100-Hz sinusoidal input to the MOAS have been reported by Sabatier.[7]

D. Measured spectra as a function of range

Because the difference between nonlinear and linear propagation is often greatest at high frequencies that are not readily visible in time waveform comparisons, the emphasis hereafter will be on trends in the measured and predicted spectra at these frequencies. The measured spectra for the 300-Hz test waveforms shown previously in Fig. 3 along 1.4° are displayed in Fig. 4. For purposes of clarity, only the levels at each of the harmonic frequencies, rather than full spectra, are displayed. The overall sound pressure level (OASPL) for each spectrum is located in the figure legend. Approximately 11 s of data were used to calculate the PSD for each case, both to allow fine-scale spectral resolution and to mitigate the effect of fluctuations in waveform amplitudes caused by wind variability. The measured harmonic levels for the same 300-Hz sine wave test, but along 0.7°, are shown in Fig. 5. Similar spectral results are obtained for both propagation angles at most distances, but there are some notable differences in the rates of high-frequency spectral decay of the 1092-m spectra. This greater variability, which is
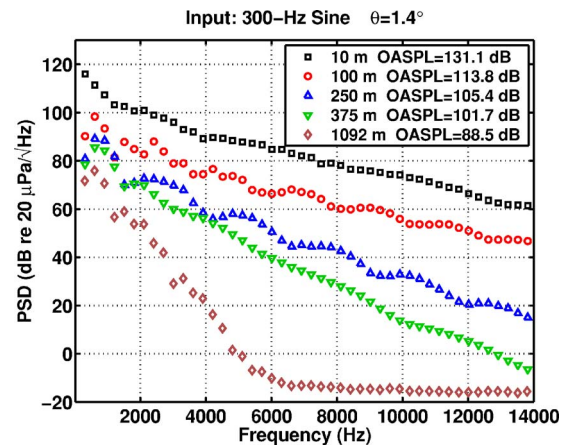
seen in the majority of the 1092-m measurements, directly impacts the ability to model the propagation out to that distance.

Some discussion of observed ground effects in the spectra is also merited. The broadband spectral results of a propagation measurement with a pink noise input are helpful in this regard. Displayed in Fig. 6 are measured spectra for a pink noise along 0.7°. A comparison of these spectra with those from the 300-Hz sine wave test in Fig. 5 reveal the presence of several spectral minima that occur at similar frequencies. In Fig. 6, the lowest frequencies at which a spectral minimum occurs for each curve (e.g., approximately 85 Hz at 10 m and 250 Hz at 100 m) closely matches those predicted by a ground reflection model using flow resistivity values corresponding to soft terrain. The particular ground interaction model used accounts for the interaction of spherically spreading waves with a finite-impedance ground[8] as well as the effects of atmospheric turbulence.[9] The results of the ground reflection model and the consistency of the measured frequencies of the lowest spectral minima indicate that these minima are due to the superposition of the direct and ground-reflected waveforms at the microphone. At higher frequencies, spectral minima are not nearly as defined. This
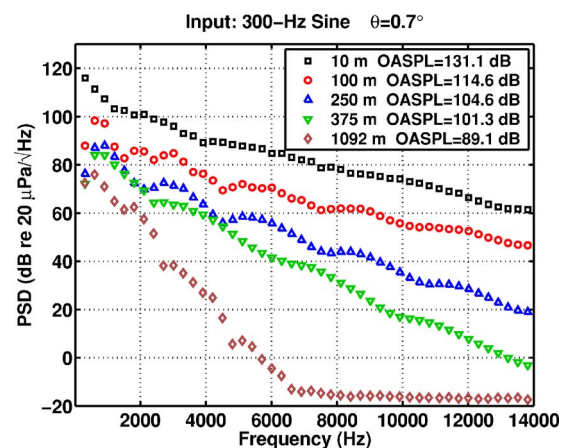
FIG. 5. (Color online) Measured MOAS harmonic levels along 0.7° for a 300-Hz input sine wave.

Gee *et al.*: Nonlinear propagation of periodic signals
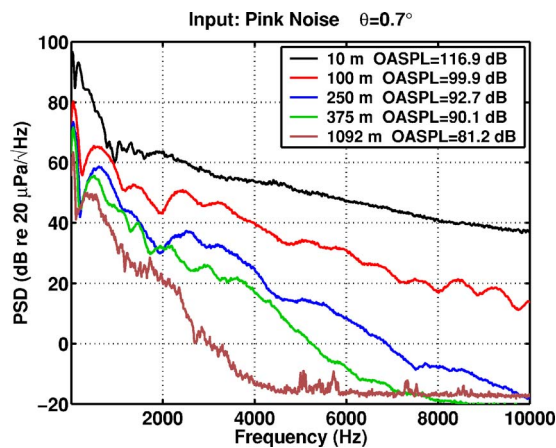
FIG. 6. (Color online) Measured MOAS spectra along 0.7° for input pink noise.

FIG. 7. (Color online) Measured MOAS harmonic levels at three microphone heights at a range of 375 m for a 300-Hz input sine wave. The time of the test was 13:34 EST.

minima broadening is likely caused by atmospheric turbulence and is predicted by the ground reflection model when turbulence is included. Because the frequencies at which these minima occur also vary somewhat from test to test, other atmospheric inhomogeneities and curved ray propagation likely influence some of the measurements.

Spectral calculations as a function of range have been made for many more cases; however, the spectra shown are sufficient to highlight important aspects of the measurements obtained, as well as to point out the measurement system noise floor. Consideration of the system noise floor is particularly important in analyses of the 375-m and 1092-m data. From the preceding figures, it is evident that for the microphones mounted on the pole at 375 m, the system noise floor was approximately −20 dB re 20 μPa/$\sqrt{\text{Hz}}$. For the 1092-m data, the noise floor was about −15 dB re 20 μPa/$\sqrt{\text{Hz}}$. Given these noise floors, the upper frequency limit of all analyses presented hereafter has been limited to where the signal-to-noise ratio is approximately 10 dB.

E. Measured spectra as a function of height

One other important aspect in describing the measurement is to consider the stability and quiescence of the atmosphere, because atmospheric homogeneity is assumed in the numerical model. Unfortunately, because detailed meteorological data at the measurement site are not available, an alternate means of examining the local atmospheric conditions has been used. The effect of the atmosphere on acoustic propagation may be studied by comparing spectra as a function of microphone height at the same range. For a perfectly still, homogeneous atmosphere, there should be general agreement between harmonic levels for the 0.7° and in 1.4° microphones at a given range. There will be differences in individual harmonics, caused by the differences in the direct and ground-reflected paths as a function of height, but the overall trend in spectral decay as a function of frequency should remain the same. Variation from a common trend for different microphone heights indicates some form of atmospheric variability.

Harmonic levels at 375 m are shown as a function of height for two cases: the 300-Hz sine wave test previously
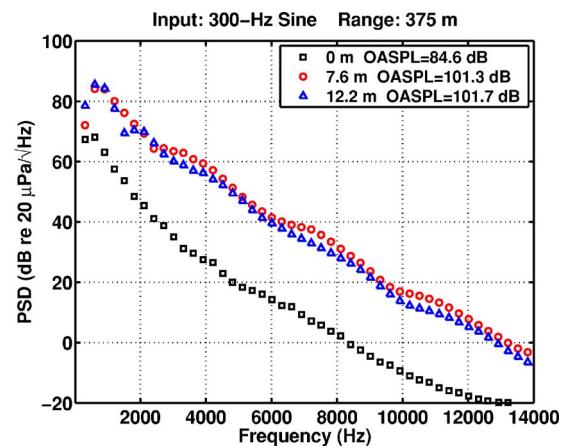
discussed, which occurred at 13:34 EST, and a 400-Hz square wave test conducted at 14:33 EST. The results for these two tests typify the change in ambient conditions that occurred during the latter measurements. Displayed in Figs. 7 and 8 are harmonic spectral levels as a function of microphone height at 375 m for the 300-Hz sine wave and 400-Hz square wave tests, respectively. As may be anticipated, the microphones located on the ground in both cases yield lower levels than for those microphones located off the ground. This is likely due to the additional losses caused by propagation along the ground at grazing incidence; consequently, ground microphone data are not considered further. For the microphones located off the ground, however, there is a significantly different behavior between the two tests with regard to the harmonic amplitudes at 7.6 m (0.7°) and 12.2 m (1.4°). For the 300-Hz test, the 0.7° levels are only slightly greater than the 1.4° levels at high frequencies. However, for the 400-Hz square wave test, the differences are much greater and increase as a function of frequency. This comparison implies that an assumption of a quiescent, homogeneous atmosphere is less appropriate for measurements taken toward the end of the day. This apparent degra-
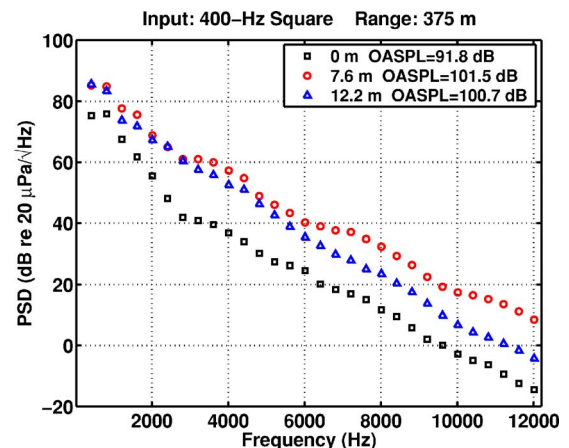
FIG. 8. (Color online) Measured MOAS harmonic levels at three microphone heights at a range of 375 m for a 400-Hz input square wave test made later in the afternoon, at 14:33 EST.

Gee *et al.*: Nonlinear propagation of periodic signals

dation of stable atmospheric conditions for later tests directly impacts accompanying propagation predictions made using the numerical model described below.

Although this summary of the complete MOAS field experiment has been carried out with relatively few cases, the waveforms and spectra shown are representative of the entire test. After a brief overview of the numerical model used to obtain nonlinear and linear propagation predictions, comparisons between measured and predicted spectra for various cases are presented and discussed.

III. NUMERICAL MODEL OVERVIEW

A. The generalized Burgers equation

In this section, the nonlinear numerical model used in the comparisons against the MOAS field experiment is described. A parabolic model equation that has been used extensively to treat the problem of nonlinear propagation through a lossy medium is the Burgers equation. In its most basic form, the Burgers equation describes plane-wave propagation through a thermoviscous medium; however, other formulations have incorporated geometrical spreading[10] and arbitrary absorption and dispersion.[11] One form of the generalized Burgers equation (GBE), on which the numerical model is based, may be written for assumed spherical spreading as

$$\frac{\partial p}{\partial r} = \frac{\beta}{2\rho_0 c_0^3} \frac{\partial p^2}{\partial \tau} + \psi(\tau)p - \frac{1}{r}p. \tag{1}$$

In Eq. (1), $p(r,\tau)$ is the acoustic pressure, r, the range variable, β, the coefficient of nonlinearity, ρ_0, the ambient density, c_0, the small-signal sound speed, $\tau = t - (r - r_0)/c_0$, the retarded time of propagation between r_0 and r, and $\psi(\tau)$, a generalized absorption and dispersion operator that acts on p. In the context of the current propagation problem, $\psi(\tau)$ represents atmospheric absorption and dispersion. An additional term may be included in the GBE to treat atmospheric stratification (e.g., see Ref. 12), but because atmospheric inhomogeneity is not modeled in this work, the term has been neglected.

B. GBE solution technique

The solution technique to the GBE employed in this research is an adaptation from previous work carried out at the University of Texas at Austin and the University of Mississippi, where hybrid time-frequency domain nonlinear propagation algorithms have been developed and refined. The time-frequency domain solution method originated with Pestorius and Blackstock,[6] who investigated finite-amplitude noise propagation in a one-dimensional tube and developed an algorithm to numerically propagate acoustic pressure waveforms. The nonlinear portion of the propagation was carried out in the time domain, and the small-signal portion of the propagation (boundary-layer absorption and dispersion) was handled in the frequency domain. Pierce[13] has demonstrated that the "Pestorius algorithm" reduces to the GBE appropriate for plane-wave propagation in a one-dimensional duct.

Anderson[14] developed an alternative to the Pestorius algorithm in his study of spherically decaying N waves. To eliminate the need for weak shock theory, which Pestorius' method employed, adaptive step sizing was used to ensure that the waveform was singly valued for all range steps. Following Anderson's initial work, improvements to the "Anderson algorithm" were made by others also interested in the nonlinear propagation of transients.[15–17]

The present numerical solution to the GBE in Eq. (1) has been developed from a study of the Pestorius and Anderson methodologies. The algorithm most closely resembles the Anderson approach because it employs an adaptive step size. However, because Anderson's code and its subsequent modifications have been intended primarily for the propagation of transients, some elements of the model are more closely linked to the work of Pestorius, who dealt with continuous-wave signals. Implementation details for the model may be found in Ref. 5.

IV. COMPARISON RESULTS AND ANALYSIS

In this section, comparisons are made between measured spectra and predicted spectra obtained by numerically propagating recorded waveforms both linearly and nonlinearly. The linear predictions, which have been obtained by removing the nonlinear term from Eq. (1), are equivalent to free-field extrapolations of the input spectrum using atmospheric absorption and spherical spreading. Use of approximately 11 s of data (2^{20} samples) in the numerical propagation constitutes a significant difference from Theobald's study, in which only a single cycle of the recorded waveform was used.[1] As with the PSD calculations in Sec. II, the use of relatively long waveforms as inputs to the numerical models allows for both fine-scale resolution and a large number of ensemble averages when calculating a predicted PSD. Note that a long input waveform and a greater number of averages are not expected to reduce errors between measured and predicted spectra if atmospheric and ground effects cause the actual propagation path to significantly differ from the straight numerical propagation path. An increased waveform length, however, has been found to help minimize variability in predicted spectra caused by wind-induced waveform amplitude fluctuations. Finally, data collected at 100 m, rather than at 10 m, have been used as inputs to the model because the 10-m on-axis microphone was found to be located within the geometrical near field of the MOAS, where an assumption of spherical spreading is not valid.[5]

In order to quantitatively compare the results of the nonlinear and linear models, an assessment of the overall error of a given model relative to the measurement is needed. The metric that has been selected for the purpose of these comparisons is the mean absolute error, in dB, between the measured PSD and a predicted PSD. For example, the mean absolute error, $\overline{E_{M,N}}$, between the measured PSD (PSD_M) and nonlinearly predicted PSD (PSD_N) may be written as

$$\overline{E_{M,N}} = \langle |PSD_M - PSD_N| \rangle, \tag{2}$$

where $\langle \ \rangle$ is the expectation operator and the spectral densities are calculated in dB. The mean absolute error between
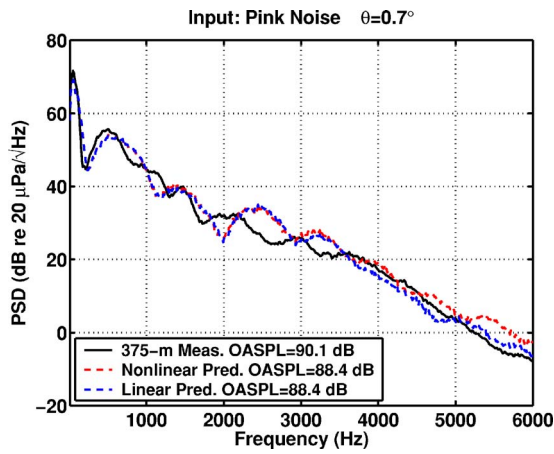
FIG. 9. (Color online) Measured and predicted spectra at 375 m, $\theta = 0.7°$, for pink noise input. The input for the predictions was the measured waveform at 100 m.

FIG. 10. (Color online) Measured and predicted harmonic levels at 375 m, $\theta = 0.7°$ for a 300-Hz input sine wave. The input for the predictions was the measured waveform at 100 m.

the measurement and linear prediction, $\overline{E_{M,L}}$, is calculated by replacing PSD_N in Eq. (2) with the linearly predicted PSD (PSD_L). In each of the comparisons between measurement and model, $\overline{E_{M,N}}$ and $\overline{E_{M,L}}$ are calculated over the frequency range where the measured levels are at least 10 dB above the system noise floor.

The comparisons that follow are broken into two broad categories. First, results from the pink noise test are analyzed because its relatively low OASPL at 100 m make the test suitable as a sort of linear benchmark between the real-world data and the simplified models. Next, comparisons between the results from various tests with periodic waveforms are made and analyzed.

A. Pink noise comparison

The OASPL of the pink noise spectrum at 100 m was 99.9 dB *re* 20 μPa (see Fig. 6), which is significantly lower than the tests with periodic waveforms. This case is useful in determining the level of agreement between the models and the measurement that is achieved when nonlinear effects appear to be minimal and when the atmosphere is relatively homogeneous. For this comparison, the recorded 100-m waveform at 0.7° has been propagated out to 375 m.

Shown in Fig. 9 is the measured 375-m PSD for 0.7°, along with the nonlinearly and linearly predicted spectra calculated from numerical propagation from 100 m. The nonlinear and linear predictions differ noticeably above 3.5 kHz, but result in the same predicted OASPL (88.4 dB *re* 20 μPa) and generally follow the decay of the measured PSD out to 6 kHz. For this case, $\overline{E_{M,N}}$ and $\overline{E_{M,L}}$ are, respectively, 2.9 and 2.6 dB, which means that the linear prediction has slightly less error than the nonlinear prediction but both models have less than 3 dB of mean absolute error between 0 and 6 kHz. The maximum difference between both models and the measurement occurs at about 2.5 kHz, where a relative spectral maximum at 100 m is propagated outward with a free-field assumption and is then compared to a measured relative spectral minimum at 375 m. This maximum error is likely due to differences in multipath interference effects at higher frequencies. This pink noise comparison indicates that the
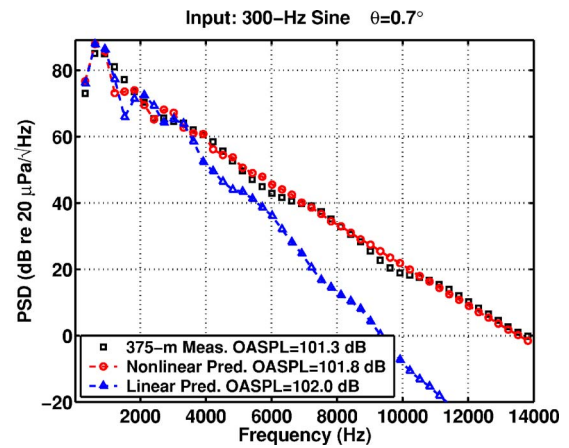
linear and nonlinear free-field models are useful for comparing with general trends observed in the measured spectra when the atmosphere is reasonably homogeneous. It also indicates, however, that agreement at specific frequencies may be poor because of the variation in multipath interference as a function of range.

B. Periodic signal comparisons

1. 100–375-m comparisons

The first set of comparisons with periodic source waveforms is carried out between 100 and 375 m, as was done with the pink noise test. The fact that virtually all periodic waveform measurements made over this range reveal a significant difference between linearly predicted and measured spectra at high frequencies suggests that nonlinearity influences the propagation. The results of comparisons between the measurement and the numerical predictions for three cases are now shown: the previously discussed 300-Hz sine-wave test along both 0.7° and 1.4° (see Figs. 4 and 5), as well as a 400-Hz sine-wave test along 0.7°. The 300-Hz measurement and predictions at 375 m along 0.7° and 1.4° are, respectively, shown in Figs. 10 and 11. The 400-Hz results along 0.7° are displayed in Fig. 12. Again, only the levels at each of the harmonics are shown for purposes of clarity. Calculations of $\overline{E_{M,L}}$ and $\overline{E_{M,N}}$ are shown in Table I for each of the three tests. Both the graphical and tabulated results reveal that the nonlinearly predicted spectra match the measurements significantly better than the linear predictions. As with the pink noise comparison, there are significant discrepancies at individual frequencies, but $\overline{E_{M,N}} < 3$ dB for each of these cases. These comparisons indicate that nonlinear propagation is the primary cause of the discrepancy between linearly predicted and measured spectral levels at high frequencies.

2. 100–1092-m comparisons

The results for the 100-375-m comparisons confirm that nonlinear effects are present in the propagation of periodic signals from the MOAS. Another point of discussion is the influence of nonlinearity in the propagation beyond 375 m
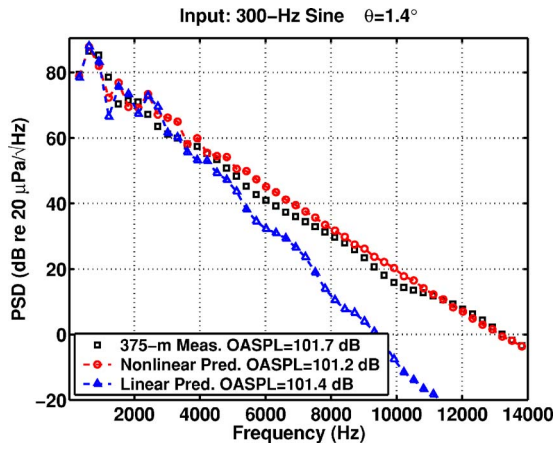
FIG. 11. (Color online) Measured and predicted harmonic levels at 375 m, $\theta=1.4°$ for a 300-Hz input sine wave. The input for the predictions was the measured waveform at 100 m.

out to 1092 m. The 300-Hz sine-wave test discussed previously was selected for study because of the good agreement achieved between the nonlinear model and measurement between 100 and 375 m. The 100-m waveforms at 0.7° and 1.4° were again used as inputs and this time were propagated out to 1092 m using the nonlinear and linear models. The results of the spectral comparisons are displayed in Figs. 13 and 14 and the mean absolute errors calculated in Table II. Although these results are not shown, nearly identical graphical and quantitative results were obtained when the 375-m recorded waveforms, rather than the 100-m recorded waveforms, were used as inputs to the model. In both cases considered here, as well as for the majority of other measurements performed, the nonlinearly predicted harmonic levels at 1092 m are consistently greater than the measured levels and yield a greater mean absolute error than the 100−375-m comparisons. For the 0.7° propagation angle, $\overline{E_{M,L}}$ is still significantly greater than $\overline{E_{M,N}}$, but for 1.4°, the two error calculations are approximately equal.

The trends seen in the 1.4° results for the 300-Hz test are common for several different tests at one or sometimes both propagation angles in that $\overline{E_{M,L}}$ and $\overline{E_{M,N}}$ are nearly
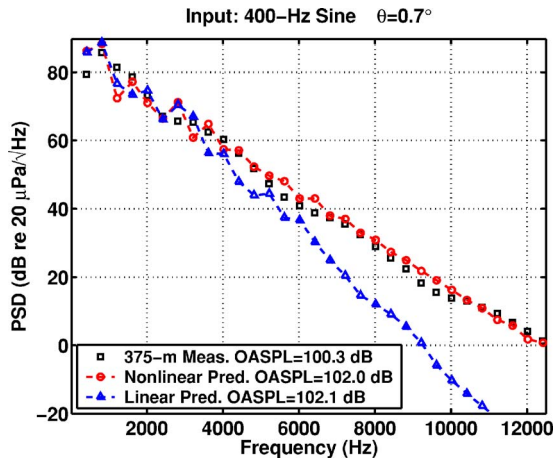
FIG. 12. (Color online) Measured and predicted harmonic levels at 375 m, $\theta=0.7°$ for a 400-Hz input sine wave. The input for the predictions was the measured waveform at 100 m.

TABLE I. Linear and nonlinear mean absolute errors, in dB, for the 100−375-m comparisons in Figs. 10−12. See Eq. (2) and the accompanying text for the definitions of $\overline{E_{M,L}}$ and $\overline{E_{M,N}}$.

Frequency/angle	$\overline{E_{M,L}}$	$\overline{E_{M,N}}$
300-Hz/0.7°	17.9	1.6
300-Hz/1.4°	16.2	2.6
400-Hz/0.7°	13.6	2.5
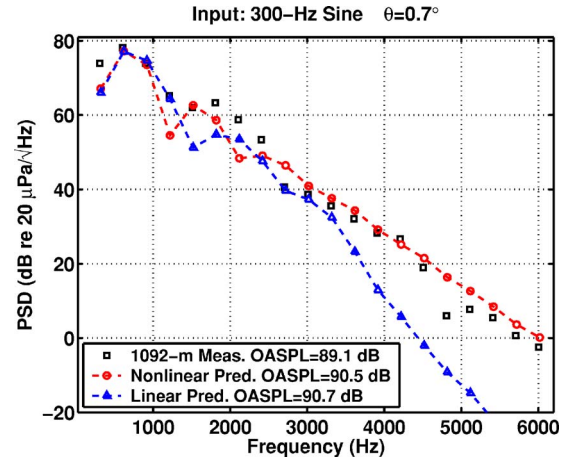
FIG. 13. (Color online) Measured and predicted harmonic levels at 1092 m, $\theta=0.7°$ for a 300-Hz input sine wave. The input for the predictions was the measured waveform at 100 m.
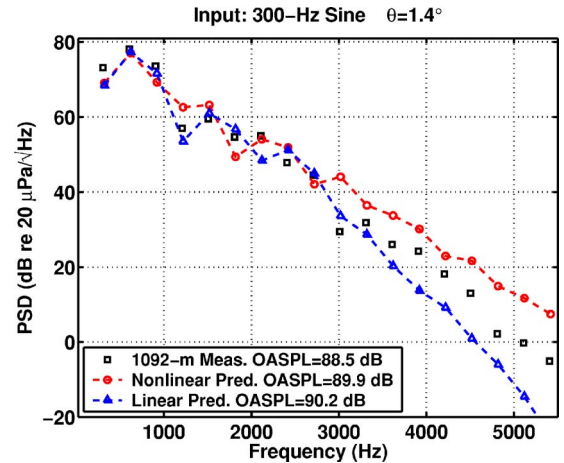
FIG. 14. (Color online) Measured and predicted harmonic levels at 375 m, $\theta=1.4°$ for a 300-Hz input sine wave. The input for the predictions was the measured waveform at 100 m.

TABLE II. Linear and nonlinear mean absolute errors, in dB, for the 100−1092-m comparisons in Figs. 13 and 14. See Eq. (2) and accompanying text for the definitions of $\overline{E_{M,L}}$ and $\overline{E_{M,N}}$.

Frequency/angle	$\overline{E_{M,L}}$	$\overline{E_{M,N}}$
300-Hz/0.7°	12.2	3.9
300-Hz/1.4°	6.2	6.4

equal and the measured spectrum falls between the nonlinear and linear predictions. Although the exact cause of the general increased disagreement between nonlinearly predicted and measured spectra over the larger propagation range has not been ascertained, some qualitative discussion is merited. Generally good agreement between the 300-Hz measurement and the nonlinear model was achieved for the 100–375-m comparisons, despite neglecting phenomena such as ground effects, turbulence, and a nonuniform atmosphere. Over a significantly longer range, each phenomenon could more significantly affect the rate of nonlinear energy transfer in the propagating waveform. As an example, the effects of turbulence, while often negligible over shorter distances, would certainly play a larger role over an increased propagation range. The results of a study of the average effect of turbulence on shock-wave rise times[18] have indicated that a steepened waveform passing through turbulence unsteepens more quickly than for passage though a quiescent medium. A more rapid unsteepening could account for the experimentally observed accelerated reduction in the high-frequency sound pressure levels. Also, a homogeneous, quiescent atmosphere and straight-ray propagation have been assumed in the numerical calculations, whereas the influence of a realistic sound-speed profile and curved rays could substantially affect the measured results at a given microphone over a longer range. For example, upwind propagation could cause ray tube divergence, thereby reducing sound amplitudes along a given ray and slowing the rate of nonlinear distortion.

3. Comparisons for later measurements

Another comparison that shows the potential for meteorological affects to influence the ability to measure nonlinearity may be made with the 400-Hz square wave measurement made later in the afternoon. It was shown previously in the discussion of measured spectra as a function of microphone height that this measurement exhibited increased dependence of level on height at high frequencies relative to earlier tests (cf. Figs. 7 and 8). In conjunction with this apparent decrease in atmospheric homogeneity, concordance between nonlinear model and measurement significantly
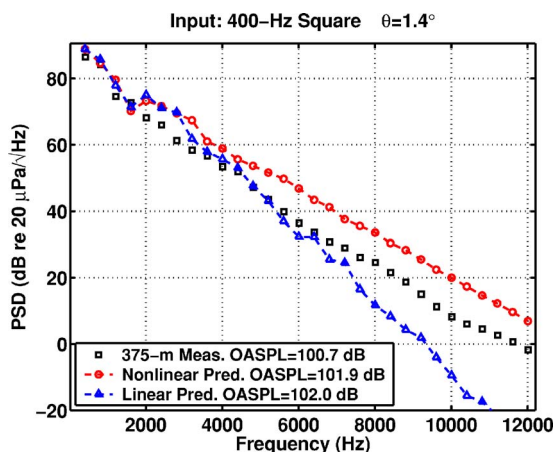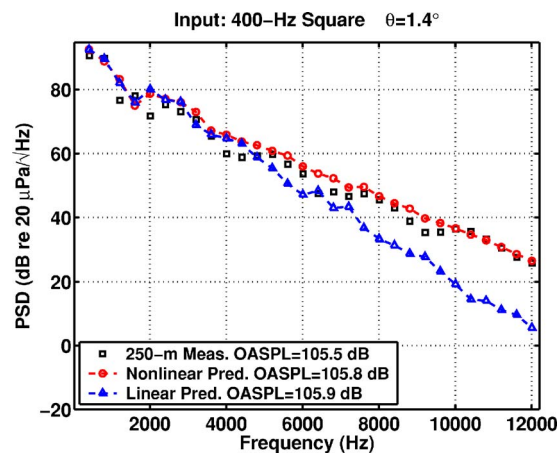
FIG. 16. (Color online) Measured and predicted harmonic levels at 250 m, $\theta=1.4°$ for a 400-Hz input square wave. The input for the predictions was the measured waveform at 100 m.

lessens. Figure 15 compares the 1.4° measured harmonic levels at 375 m with those predicted from numerical propagation from 100 m for the 400-Hz square wave test. The 100-m-based numerical calculation significantly overpredicts the measured sound pressure levels at high frequencies. To further investigate the cause of this discrepancy, the numerical propagation path was divided up using the intermediate measurement at 250 m (cf. Fig. 2) and separate calculations were performed for the 100–250-m and 250–375-m ranges. The results of these comparisons are shown in Figs. 16 and 17, respectively. Although the algorithm continues to slightly overpredict the spectral levels between 250 and 375 m, agreement is substantially better over the shorter propagation distances. The agreement is quantified with calculations of $\overline{E_{M,L}}$ and $\overline{E_{M,N}}$ in Table III. The fact that significantly better agreement is achieved with these shorter-range calculations than with the full 100–375-m range indicates that atmospheric effects can greatly impact conclusions regarding the nonlinearity of the propagation, even over relatively short distances.

FIG. 15. (Color online) Measured and predicted harmonic levels at 375 m, $\theta=1.4°$ for a 400-Hz input square wave. The input for the predictions was the measured waveform at 100 m.
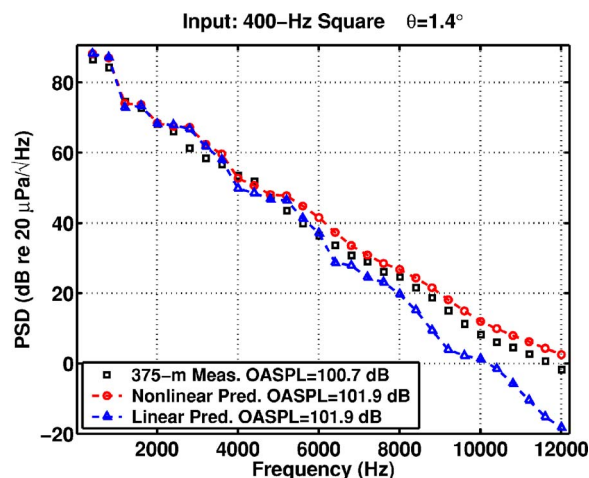
FIG. 17. (Color online) Measured and predicted harmonic levels at 375 m, $\theta=1.4°$ for a 400-Hz input square wave. The input for the predictions was the measured waveform at 250 m.

TABLE III. Linear and nonlinear mean absolute errors, in dB, for the 400-Hz square-wave test comparisons in Figs. 15–17. See Eq. (2) and the accompanying text for the definitions of $\overline{E_{M,L}}$ and $\overline{E_{M,N}}$.

Range	$\overline{E_{M,L}}$	$\overline{E_{M,N}}$
100–375 m	9.1	7.8
100–250 m	8.0	2.7
250–375 m	5.3	2.9

V. CONCLUSIONS

Finite-amplitude propagation effects have been measured in the outdoor propagation of periodic signals generated using the U.S. Army Research Laboratory's Mobile Acoustic Source (MOAS). For most cases, the results of the nonlinear model based on the generalized Burgers equation compare favorably with the measurement and demonstrate significantly less mean absolute error than results from a linear model. This is true despite the fact that the nonlinear model assumes free-field propagation through a homogenous atmosphere and therefore neglects some of the conditions of the actual experiment. The results of the comparisons indicate that effects of nonlinear propagation are readily observable between 100 and 375 m, despite overall sound pressure levels (OASPL) at 100 m that are less than 115 dB re 20 μPa for all measurements (see the legend in Fig. 5, where the maximum OASPL is 114.6 dB re 20 μPa). Based on OASPL alone, these results indicate that nonlinearity is very likely to occur in high-amplitude jet noise propagation, for which levels at 100 m may be 10–15 dB greater than for the MOAS. This corroborates earlier statements of Webster and Blackstock,[2] who reached a similar conclusion based on the results of their relatively short-range propagation experiments. Other comparisons shown in the present work are also important, specifically for the cases of propagation over greater distances and measurements made later in the afternoon, where agreement between the nonlinear model and experiment are not as good. They point to the need for the application of more sophisticated modeling techniques and additional research to better understand and predict the finite-amplitude propagation of continuous waveforms through a spatially and temporally variable atmosphere.

ACKNOWLEDGMENTS

[1]M. A. Theobald, "Experimental study of outdoor propagation of spherically spreading periodic acoustic waves of finite amplitude," Technical Report ARL-TR-77-5, Applied Research Laboratories, The University of Texas at Austin, ADA 039020, 1977.

[2]D. A. Webster and D. T. Blackstock, "Experimental investigation of outdoor propagation of finite-amplitude noise," NASA Contractor Report 2992, Applied Research Laboratories, The University of Texas at Austin, 1978.

[3]The MOAS has often been referred to as the "Mother Of All Speakers" (e.g., see Ref. 7).

[4]V. W. Sparrow, K. L. Gee, J. M. Downing, and K. J. Plotkin, "Military aircraft noise and nonlinear acoustics," J. Acoust. Soc. Am. **112**, 2214 (2002).

[5]K. L. Gee, "Prediction of nonlinear jet noise propagation," Ph.D. thesis, Graduate Program in Acoustics, The Pennsylvania State University, 2005.

[6]F. M. Pestorius and D. T. Blackstock, "Propagation of finite-amplitude noise," in *Finite-Amplitude Wave Effects in Fluids*, edited by L. Bjorno (IPC Science and Technology Press, Guildford, Surrey, England, 1973), pp. 24–29.

[7]J. M. Sabatier, "Acoustical characterization of the mother of all speakers," National Center for Physical Acoustics, The University of Mississippi, NCPA Control No.-JMS0593-1, May 1993.

[8]K. Attenborough, S. I. Hayek, and J. M. Lawther, "Propagation of sound above a porous half-space," J. Acoust. Soc. Am. **68**, 1493–1501 (1980).

[9]G. A. Daigle, "Effects of atmospheric turbulence on the interference of sound waves above a finite impedance boundary," J. Acoust. Soc. Am. **65**, 45–49 (1979).

[10]R. V. Khokhlov, K. A. Naugol'nykh, and S. I. Soluyan, "Waves of moderate amplitudes in absorbing media," Acustica **14**, 241–247 (1964).

[11]D. T. Blackstock, "Generalized Burgers equation for plane waves," J. Acoust. Soc. Am. **77**, 2050–2053 (1985).

[12]R. O. Cleveland, J. P. Chambers, H. E. Bass, R. Raspet, D. T. Blackstock, and M. F. Hamilton, "Comparison of computer codes for the propagation of sonic boom waveforms through isothermal atmospheres," J. Acoust. Soc. Am. **100**, 3017–3027 (1996).

[13]A. D. Pierce, "Progressive wave equations and algorithms for sonic boom propagation," *Proc. Noise-Con 93*, Noise Control Foundation, Poughkeepsie, New York, 1993, pp. 157–162.

[14]M. O. Anderson, "The propagation of a spherical N wave in an absorbing medium and its diffraction by a circular aperture," Technical Report ARL-TR-74-25, Applied Research Laboratories, The University of Texas at Austin, AD 787878, 1974.

[15]H. E. Bass and R. Raspet, "Vibrational relaxation effects on the atmospheric attenuation and rise times of explosion waves," J. Acoust. Soc. Am. **64**, 1208–1210 (1978).

[16]L. B. Orenstein, "The rise time of N waves produced by sparks," Technical Report ARL-TR-82-51, Applied Research Laboratories, The University of Texas at Austin, ADA 120817, 1982.

[17]H. E. Bass, J. Ezell, and R. Raspet, "Effect of vibrational relaxation on rise times of shock waves in the atmosphere," J. Acoust. Soc. Am. **74**, 1514–1517 (1983).

[18]P. Blanc-Benon, B. Lipkens, L. Dallois, M. F. Hamilton, and D. T. Blackstock, "Propagation of finite amplitude sound through turbulence: Modeling with geometrical acoustics and the parabolic approximation," J. Acoust. Soc. Am. **111**, 487–498 (2002).

Technique to minimize couplant-effect in acoustic nonlinearity measurements

Li Sun, Salis S. Kulkarni, Jan D. Achenbach, and Sridhar Krishnaswamy[a)]

Center for Quality Engineering and Failure Prevention, Northwestern University, Evanston, Illinois 60208

A modified technique to measure acoustic nonlinearity in fatigued components is proposed in this paper. The advantage of the proposed technique is that it minimizes measurement errors due to the couplant between the transducers and the specimen. Measurements are performed on a fatigued steel 4340 specimen and the coefficients of variation of the nonlinearity parameter are calculated. It is shown that the coefficients of variation of the nonlinearity parameter obtained using the proposed technique are approximately half of that obtained using the conventional technique. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354023]

I. INTRODUCTION

The nucleation and growth of a fatigue crack in a metal is generally preceded by a process of damage accumulation at the microstructural level. As the number of fatigue cycles increases, the damage accumulation involves the formation of dislocation monopoles, followed by dislocation loops and dipoles, dislocation veins, persistent slip bands, and subsequent nucleation of microcracks.[1,2] These changes in the microstructure due to damage accumulation introduce changes in the material response, and in particular lead to nonlinear stress-strain response. Therefore by measuring the changes in the material nonlinearity, the damage accumulated during the fatigue process can be estimated. To characterize the material nonlinearity, the usual way is to launch a single frequency ultrasonic wave into the specimen and receive the signal at a certain distance. Due to the material nonlinearity, the single frequency wave is distorted and second and higher harmonics are generated.[3–7] As a result, the received signal not only consists of the component at the transmitting (fundamental) frequency, but also the second and higher harmonics whose amplitudes increase as the material nonlinearity increases. Hence, the material nonlinearity in a specimen can be quantified in terms of the amplitude of the fundamental and higher harmonics generated by propagating a monochromatic ultrasonic wave. This phenomenon of distortion of an acoustic wave due to material nonlinearity is generally referred to as acoustic nonlinearity. The above discussion indicates that acoustic nonlinearity can be used as a parameter to monitor the precrack fatigue damage process.[8]

Conventional techniques to measure changes in the acoustic nonlinearity during fatigue are mostly based on measuring the amplitudes of the fundamental and the second harmonic generated using either bulk or surface waves. Experimental results demonstrating the changes of acoustic nonlinearity during fatigue process using bulk waves have been reported in papers on fatigued stainless steel 410-Cb,[9,10] titanium alloys[11] and 2024-T4 aluminum alloy.[10,12] Similar experiments done using surface acoustic waves have also been reported on fatigued 7073-aluminium,[13] titanium alloy Ti-6Al–4V and inconel 718,[14] and concrete.[15]

The most common way to make acoustic measurements is by using contact piezoelectric transducers. To facilitate signal transmission, a couplant layer is applied between the transducer and specimen. However, the couplant layer leads to variations in ultrasonic measurements, and the results are sensitive to coupling pressure, couplant layer thickness and viscosity, etc.[16] A self-calibrating ultrasonic technique for crack depth measurement using surface waves, which eliminates the couplant effect, was proposed by Achenbach *et al.*[17] Their basic idea is to perform a set of four measurements which enable the elimination of the couplant response in linear ultrasonic measurements.

Couplant variability is especially problematic for acoustic nonlinearity monitoring of fatigue damage since these measurements have to be made repeatedly as a component is fatigue cycled, and variations in measurement due to the couplant make quantitative comparisons difficult. This is true for both laboratory measurements where the transducers may be repeatedly removed during fatigue cycling and reattached periodically for acoustic nonlinearity measurements, and measurements using permanently mounted transducers where the couplant can be expected to degrade with time as the fatigue cycling progresses. In this paper, we propose a method to minimize the effect of the couplant in the acoustic nonlinearity measurements. Since we are primarily interested in minimizing the couplant effects during fatigue monitoring, we present results comparing the nonlinearity measurements based on the second harmonic generation using both the conventional method and the proposed method. The proposed method can also be extended in a straightforward manner to minimize couplant effects for nonlinearity measurements based on higher harmonic generation. This is demonstrated by comparing the nonlinearity measurements based on third harmonic generation using the conventional method and the proposed method.

[a)]Electronic mail: s-krishnaswamy@northwestern.edu

II. ACOUSTIC NONLINEARITY MEASUREMENT

For the one-dimensional problem of longitudinal wave propagation through a solid containing mobile dislocations, the amplitude of the fundamental at a distance l from the source can be approximated as[18,19]

$$A_1(l) = A_1(0)f_1(\alpha_1, l), \tag{1}$$

where $A_1(0)$ is the fundamental amplitude at the source and $f_1(\alpha_1, l)$ is a function of the attenuation coefficient α_1 at the fundamental frequency and the propagation distance l. Since we are primarily interested in monitoring the fatigue process, the effect of dislocations on the generation of higher harmonics as well as their effect on attenuation need to be considered. For longitudinal waves at a fixed distance from the source, it can be shown that the amplitude of the second harmonic is proportional to the square of the fundamental, and the amplitude of the third harmonic is proportional to the cube of the fundamental.[18–20] The amplitude of the nth harmonic at a distance l from the source can therefore be represented in the following form:

$$A_n(l) = \gamma_n[A_1(0)]^n f_n(\alpha_1, \alpha_2, \ldots, \alpha_n, l), \quad (n=2,3), \tag{2}$$

where $\gamma_n (n=2,3)$ are the parameters which depend on the fundamental frequency and material properties such as density, microstructure, and second- and third-order elastic constants. The parameters $\gamma_n (n=2,3)$ are expected to increase with the increase of the material nonlinearity during the fatigue process, and henceforth are referred to as nonlinearity parameters. In Eq. (2), $f_n(\alpha_1, \alpha_2, \ldots, \alpha_n, l), (n=2,3)$ is a factor which takes into account the attenuation and α_n is the attenuation coefficient for nth harmonic. Note that the attenuation factor $f_n(\alpha_1, \alpha_2, \ldots, \alpha_n, l), (n=2,3)$ depends on the attenuation parameters of the fundamental, second harmonic, up to the nth harmonic. Explicit expressions for the attenuation factor for the second harmonic and third harmonic are available in the literature.[18,19]

Similar to the case of longitudinal waves, an initially sinusoidal surface acoustic wave (SAW) also generates higher harmonics in a nonlinear medium.[21,22] A detailed discussion of various aspects of SAWs in nonlinear elastic media appears in Mayer's work.[23] Numerous mathematical theories of the evolution of nonlinear surface waves, which provide surface displacements and velocity profiles, have been proposed in the literature.[23–27] The theories for SAWs in nonlinear media (see for example Mayer[23]) also predict that at a fixed distance from the source, $A_2 \propto A_1^2$, $A_3 \propto A_1^3$. In what follows we assume that the attenuation effect for SAWs can be accounted for in a manner similar to that for the longitudinal waves, as expressed by Eqs. (1) and (2).

In the following, models for the conventional acoustic nonlinearity measurement technique and the proposed couplant-effect-minimization measurement technique are developed. Note that they are applicable for both longitudinal and surface acoustic waves. To develop these models, the following assumptions are made:

- The relevant components of the measurement system—amplifier, preamplifier, transducers, are linear, i.e., there is no harmonic generation in these components.
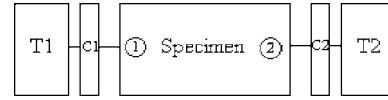
FIG. 1. Configuration for nonlinearity measurement.

- There is no harmonic generation due to the couplant layer between the transducers and the specimen (reasonable given that the layer is typically very thin).

From the above assumptions, it follows that the material is the sole contributor to the observed nonlinearity.

A block diagram of nonlinearity measurements using contact transducers is shown in Fig. 1 with transducer 1 ($T1$) located at position (1) and transducer 2 ($T2$) located at position (2). The couplant layers between the two transducers and specimen are denoted as $C1$ and $C2$, respectively.

The following notation is used in the remainder of the paper:

$D_i(\omega)$–amplitude of the input voltage to transducer i as a function of frequency ω

$G_i(\omega)$–response function of transducer i when it is used as a transmitter

$R_i(\omega)$–response function of transducer i when it is used as a receiver

$C_i(\omega)$–couplant response function at position i

$P_{ij}(\omega)$–function that takes into account the attenuation of the signal received by transducer j at frequency ω when the input signal is transmitted at frequency ω from transducer i

$\bar{P}_{ij}(\omega_1, \omega_2)$–function that takes into account the attenuation of the signal received by transducer j at frequency ω_2 when the input signal is transmitted at frequency ω_1 from transducer i

$V_{[j](\omega_2)}^{[i](\omega_1)}$–amplitude of the signal received at transducer j at frequency ω_2 when the input signal is transmitted at frequency ω_1 from transducer i.

The response function of the transducer in transmission is the frequency response of the transducer for converting the electrical energy into sound energy, while the response function of the transducer in reception is the frequency response of the transducer for converting the sound energy into electrical energy. The response function of the couplant is the frequency response of the couplant for transmitting sound energy. It takes into account the effect of attenuation and depends on the contact pressure, couplant layer thickness and viscosity. It should be noted that in the following discussions we will only consider the *amplitude* frequency response, and will not explicitly track the phase response.

A. Conventional acoustic nonlinearity measurements

In the conventional approach, one transmits a signal at frequency ω_0 through transducer 1 and measures the amplitude of the signal at frequencies ω_0 and $n\omega_0 (n=2,3)$ at transducer 2. From the above assumptions, it follows that the voltage signal received by transducer 2 at frequency ω_0 is given by

$$V^{[1](\omega_0)}_{[2](\omega_0)} = [D_1(\omega_0)G_1(\omega_0)C_1(\omega_0)]P_{12}(\omega_0)C_2(\omega_0)R_2(\omega_0). \tag{3}$$

The above is a completely linear measurement. Note that here $P_{12}(\omega_0)$ takes into account the attenuation of the acoustic wave at fundamental frequency ω_0 when it travels from position 1 to position 2 in the specimen. The term in the square bracket can be identified with the injected fundamental $A_1(0)$ and the attenuation term $P_{12}(\omega_0)$ with the function $f_1(\alpha_1,l)$ in Eq. (1).

Next we consider the voltage signals received by transducer 2 at frequency $n\omega_0$ ($n=2,3$) due to signal generated at frequency ω_0 by transducer 1. These are nonlinear contributions. As mentioned earlier, the amplitude of the nth harmonic is proportional to the nth power of the amplitude of the fundamental signal. Therefore we can write

$$V^{[1](\omega_0)}_{[2](n\omega_0)} = \overline{\gamma_n}[D_1(\omega_0)G_1(\omega_0)C_1(\omega_0)]^n \bar{P}_{12}(\omega_0,n\omega_0)$$
$$\times C_2(n\omega_0)R_2(n\omega_0), \quad (n=2,3) \tag{4}$$

where $\overline{\gamma_n}(n=2,3)$ are functions of the material nonlinearity. Note that the constant $\overline{\gamma_n}(n=2,3)$ are different from the constants $\gamma_n(n=2,3)$ introduced in Eq. (2) since we are now relating the voltage amplitudes while earlier we were relating the displacement amplitudes. As fatigue progresses, we can expect $\overline{\gamma_n}(n=2,3)$ to change and therefore they serve as indicators of the accumulated damage. Note that $\bar{P}_{12}(\omega_0,n\omega_0)$ takes into account the attenuation of the acoustic signal received by transducer 2 at frequency $n\omega_0$ when the input signal is transmitted at frequency ω_0 from transducer 1, and we can identify it with the function $f_n(\alpha_1,\alpha_2,\ldots,\alpha_n,l),(n=2,3)$ in Eq. (2).

From Eqs. (3) and (4) one can derive the following expression:

$$\hat{\gamma}_n = T_R \frac{C_2^n(\omega_0)}{C_2(n\omega_0)} \frac{V^{[1](\omega_0)}_{[2](n\omega_0)}}{(V^{[1](\omega_0)}_{[2](\omega_0)})^n} \quad (n=2,3), \tag{5}$$

where

$$\hat{\gamma}_n = \frac{\overline{\gamma_n} \cdot \bar{P}_{12}(\omega_0,n\omega_0)}{P_{12}^n(\omega_0)} \quad \text{and} \quad T_R = \frac{R_2^n(\omega_0)}{R_2(n\omega_0)}.$$

Note that $\hat{\gamma}_n$ is defined in terms of $\overline{\gamma_n}$, which depend on the material nonlinearity, and $P_{12}(\omega_0,n\omega_0)$ and $P_{12}(\omega_0)$ which are functions of the attenuation coefficients. Both the material nonlinearity and the attenuation are expected to increase with fatigue.[28,29] For cases where the change in attenuation due to fatigue is significantly less than the change in the material nonlinearity, we expect $\hat{\gamma}_n$ to increase with fatigue.

Since we are only interested in calculating the changes in $\hat{\gamma}_n$ as fatigue progresses, it is not essential to calculate the quantity T_R, which depends on the response of transducer 2 and is expected to remain constant for the entire duration of the experiment. The couplant responses at ω_0 and $n\omega_0$ may change from measurement to measurement (as fatigue cycling proceeds) because of possible couplant degradation, changes in couplant thickness or contact pressure which are difficult to control. In conventional measurements, the changes in couplant response are neglected and the changes in the material nonlinearity during fatigue are quantified by calculating γ_n^{con}, with

$$\gamma_n^{con} = \frac{V^{[1](\omega_0)}_{[2](n\omega_0)}}{(V^{[1](\omega_0)}_{[2](\omega_0)})^n}, \quad (n=2,3) \tag{6}$$

at different cycle numbers. Obviously the above measurement of acoustic nonlinearity does not take into account the couplant effect, and as such great care will need to be taken to ensure that the couplant response is identical from measurement to measurement.

B. Modified acoustic nonlinearity measurements

As pointed out in the previous section, the couplant response is usually neglected when measuring the acoustic nonlinearity. But as the measured signals are sensitive to the couplant response, and since in general it is not easy to keep the coupling constant, large variations in the ratio $V^{[1](\omega_0)}_{[2](n\omega_0)}/(V^{[1](\omega_0)}_{[2](\omega_0)})^n$ are likely to be observed. This makes comparisons of the nonlinearity parameter as a function of fatigue cycling very difficult. To minimize the couplant effects, a modified technique to measure the acoustic nonlinearity is suggested in this paper. This requires that, in addition to the measurements required by the conventional technique, two more measurements be carried out:

- Transmit a signal at frequency $n\omega_0$ through transducer 1 and measure the amplitude of the signal at $n\omega_0$ at transducer 2. Similar to Eq. (3), the received signal can be written as

$$V^{[1](n\omega_0)}_{[2](n\omega_0)} = [D_1(n\omega_0)G_1(n\omega_0)C_1(n\omega_0)]P_{12}(n\omega_0)$$
$$\times C_2(n\omega_0)R_2(n\omega_0). \tag{7}$$

- Transmit a signal at frequency ω_0 through transducer 2 and measure the amplitude of the signal at $n\omega_0$ at transducer 1. Following Eq. (4), the measured signal in terms of the response function is given by

$$V^{[2](\omega_0)}_{[1](n\omega_0)} = \overline{\gamma_n}[D_2(\omega_0)G_2(\omega_0)$$
$$\times C_2(\omega_0)]^n \bar{P}_{21}(\omega_0,n\omega_0)C_1(n\omega_0)R_1(n\omega_0). \tag{8}$$

Using Eqs. (3), (4), (7), and (8) and the fact that $\bar{P}_{12}(\omega_0,n\omega_0)=\bar{P}_{21}(\omega_0,n\omega_0)$, we can show that

$$\tilde{\gamma}_n = T_{GR} \left[\frac{V^{[1](\omega_0)}_{[2](n\omega_0)} V^{[2](\omega_0)}_{[1](n\omega_0)}}{(V^{[1](\omega_0)}_{[2](\omega_0)})^n V^{[1](n\omega_0)}_{[2](n\omega_0)}} \right]^{1/2} \left[\frac{D_1(n\omega_0)}{D_2^n(\omega_0)} \right]^{1/2},$$
$$(n=2,3), \tag{9}$$

where

$$\tilde{\gamma}_n = \frac{\overline{\gamma_n} \cdot \bar{P}_{12}(\omega_0,n\omega_0)}{[P_{12}^n(\omega_0)P_{12}(n\omega_0)]^{1/2}} \quad \text{and}$$

$$T_{GR} = \left[\frac{R_2^n(\omega_0)}{G_2^n(\omega_0)} \frac{G_1(n\omega_0)}{R_1(n\omega_0)} \right]^{1/2}.$$

It is seen from Eq. (9) that the couplant response does not appear in the expression for $\tilde{\gamma}_n$, i.e., the couplant effect has been *theoretically* eliminated. As pointed out earlier, it is not necessary to measure T_{GR} (dependent on instrumentation only) since it is expected to remain constant throughout the entire duration of the experiment. Therefore, in the modified technique, the changes in the material nonlinearity during fatigue can be quantified by calculating γ_n^{mod}, where

$$\gamma_n^{\mathrm{mod}} = \left[\frac{V_{[2](n\omega_0)}^{[1](\omega_0)} V_{[1](n\omega_0)}^{[2](\omega_0)}}{(V_{[2](\omega_0)}^{[1](\omega_0)})^n V_{[2](n\omega_0)}^{[1](n\omega_0)}} \right]^{1/2} \left[\frac{D_1(n\omega_0)}{D_2^n(\omega_0)} \right]^{1/2},$$

$$(n = 2,3) \qquad (10)$$

at different cycle numbers. Observe that γ_n^{mod} is a quantity which depends on the material nonlinearity and attenuation. As mentioned earlier, both the material nonlinearity and the attenuation are expected to increase with fatigue. If the changes in the attenuation are significantly less than the changes in the material nonlinearity, we can expect γ_n^{mod} to increase with fatigue.

It should also be pointed out that the measured values of γ_n^{mod} and γ_n^{con} are expected to be quantitatively different as their expressions involve different combinations of the material nonlinearity and attenuation parameters.

III. RESULTS AND DISCUSSION

In the present work, the nonlinearity measurements are preformed for surface acoustic waves. There are two reasons for using SAWs instead of bulk waves: (a) they are easy to generate on structural components of general shape, (b) the energy density associated with SAWs is confined to the surface in a narrow spatial region, where most damage is generally accumulated during fatigue. Experimental results are presented for both the conventional and modified techniques.

The measurement system consists of a 5 MHz broadband piezoelectric transducer located at position 1 (or 2) and a 10 MHz piezoelectric transducer broadband located at position 2 (or 1) (see Fig. 1). Note that these transducers have quite a broad (but not flat) response bandwidth about their center frequencies. The distance between the two transducers was 1.8 cm. The generating signals at 5, 10, and 15 MHz were 20-cycle Hanning window-modulated tonebursts. Measurements were made on a dogbone steel 4340 specimen with thickness=2.75 mm, width=15.0 mm, and length =190 mm. The yield strength of the steel specimen was measured to be 1200 MPa, the ultimate tensile strength 1300 MPa, and the Young's Modulus 205 GPa. The specimen was subjected to a stress controlled uniaxial sinusoidal load (frequency 10 Hz) for 100 000 cycles in an MTS fatigue
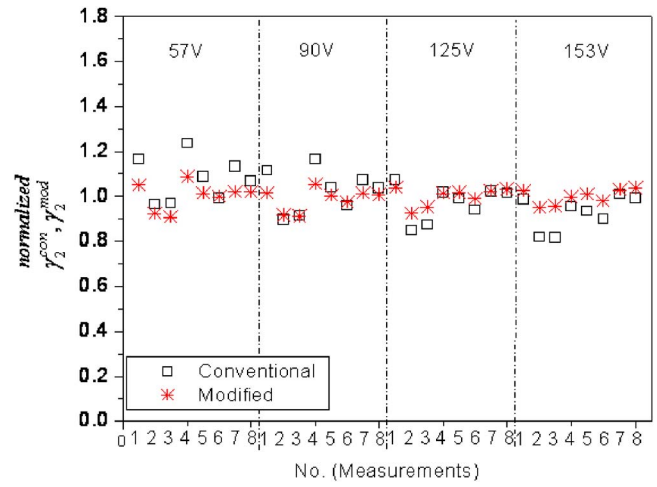
FIG. 2. Graphical comparison of the second harmonic nonlinearity measurements obtained using the conventional technique and the modified technique. The comparisons are presented for four different input signals: 57, 90, 125, and 153 V.

machine. The stress cycling was as follows: σ_{max} was set to be 950 MPa while σ_{min} was set to be 95 MPa ($R = \sigma_{\mathrm{min}}/\sigma_{\mathrm{max}} = 0.1$, $\sigma_{\mathrm{mean}} = 500$ MPa).

It is not the purpose of this paper to discuss the evolution of the acoustic nonlinearity parameter as a function of fatigue cycling (that will be reported elsewhere). Rather, the purpose here is to show that the proposed method provides a more robust measurement of the acoustic nonlinearity than conventional methods. To establish this, acoustic nonlinearity measurements were carried out after the specimen was removed from the fatigue machine (after 100 000 cycles) using both the conventional method and the modified method. Several sets of measurements were made to obtain some statistics on the scatter in the experimental measurements due to variability in the couplant. Each set of measurements was made by removing the transducers and placing them again on the specimen, not necessarily with the same contact pressure, and each time a fresh couplant layer was applied.

Second Harmonic Measurements. The second harmonic nonlinearity measurements using the conventional and modified techniques are presented. The measurements were repeated eight times and for each set of measurements four different transducer drive voltages were used: $V1 = 57$ V, $V2 = 90$ V, $V3 = 125$ V, and $V4 = 153$ V. The normalized nonlinearity parameter obtained using both the conventional and modified techniques is shown in Fig. 2.

For the case of conventional technique, each γ_2^{con} is normalized by its mean value, and for the case of modified technique, each γ_2^{mod} is normalized by its mean value. It is

TABLE I. Second harmonic nonlinearity measurements at four different input signals.

| Input | 57 V | | 90 V | | 125 V | | 153 V | |
	γ_2^{con}	γ_2^{mod}	γ_2^{con}	γ_2^{mod}	γ_2^{con}	γ_2^{mod}	γ_2^{con}	γ_2^{mod}
Mean	18.48	18.44	17.59	18.14	16.68	18.37	15.90	18.35
Standard deviation	1.69	1.10	1.64	0.92	1.34	0.75	1.30	0.62
COV	0.0916	0.0597	0.0934	0.0508	0.0804	0.0410	0.0819	0.0338
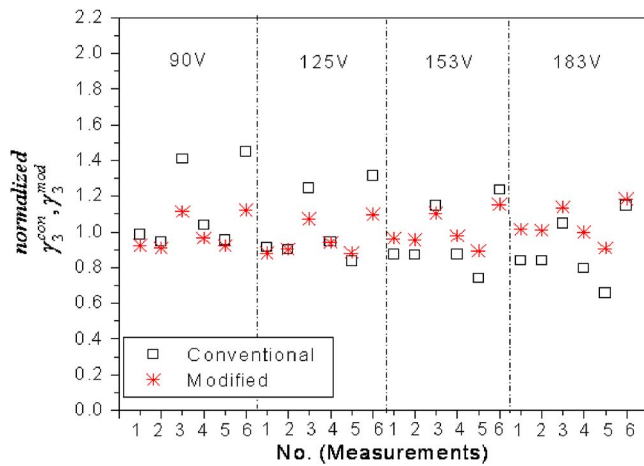
FIG. 3. Graphical comparison of the third harmonic nonlinearity measurements obtained using the conventional technique and the modified technique. The comparisons are presented for four different input signals: 90, 125, 153, and 183 V.

seen from the figure that the scatter in the nonlinearity measurements obtained using the modified technique is less than that obtained using the conventional technique. This observation is quantified by presenting the results for nonlinearity measurements and the corresponding statistical quantities in Table I.

The coefficient of variation (COV), defined as the ratio of the standard deviation to the mean value, is used as an error estimator to compare the conventional technique with the modified technique. The advantage of this error estimator as compared to the standard deviation is that it is a dimensionless, scale-invariant quantity. It is seen that for all the four different input signals, the COV of γ_2^{mod}—the parameter quantifying the scatter in the material nonlinearity following the modified technique is approximately half of that of γ_2^{con}—the parameter quantifying the scatter in the material nonlinearity following the conventional technique. While the decrease in scatter is significant, we believe that the primary cause of the residual scatter in the modified measurements is attributable to instrumentation limitations. To see this, we separately analyzed the COV for each set of experimental measurements—$V_{[2](\omega_0)}^{[1](\omega_0)}$, $V_{[2](2\omega_0)}^{[1](\omega_0)}$, $V_{[1](2\omega_0)}^{[2](\omega_0)}$, and $V_{[2](2\omega_0)}^{[1](2\omega_0)}$. The corresponding values are 0.079, 0.087, 0.096, and 0.084, respectively. Note that the COV for $V_{[1](2\omega_0)}^{[2](\omega_0)}$ is the largest. This is the least optimal measurement because it corresponds to the case where transducer 2 (central frequency at 10 MHz) is driven at 5 MHz and transducer 1 (central frequency at 5 MHz) is used to receive the second harmonic (10 MHz) signal. Therefore both the generation and the reception are

nonoptimal. To overcome this limitation, broadband transducers with flat responses and with negligible intrinsic harmonic generation need to be used.

Third Harmonic Measurements. The experimental results and corresponding statistical quantities for the third harmonic nonlinearity measurements are shown in Fig. 3 and Table II, respectively.

The measurements were repeated six times, and for each set of measurements four different transducer drive voltages were used: $V1 = 90$ V, $V2 = 125$ V, $V3 = 153$ V, and $V4 = 183$ V. To avoid the much higher second harmonic frequency components from saturating the receiver and amplifier, a 15 MHz high-pass filter and an additional 20 dB amplification in the receiver were employed in the third harmonic measurements.

Similar to the results obtained for the second harmonic case, the COV of γ_3^{mod} is approximately half of that of γ_3^{con}. This again demonstrates that the modified technique will help to reduce the variability and therefore increase the robustness in the acoustic nonlinearity measurements based on higher harmonic generation. It is worth reiterating that the measured values of γ_n^{mod} and γ_n^{con} are quantitatively different as their expressions involve different combinations of the material nonlinearity and attenuation parameters. The same is true for the parameters γ_2^{mod} and γ_3^{mod} which also involve different measures of the acoustic nonlinearity in combination with material attenuation, and as such it is not possible to directly compare the two quantities.

IV. CONCLUSIONS

The main points of this paper are:

(1) Acoustic nonlinearity measurements can be useful for monitoring fatigue damage in metallic components. Conventional acoustic nonlinearity measurements are susceptible to large scatter due to variations in coupling between the transducers and the specimens. This is especially problematic when acoustic nonlinearity measurements need to be made at different cycles for fatigue damage monitoring.

(2) A modified method for acoustic nonlinearity measurements is proposed that minimizes the couplant effect. It is shown that the scatter in the results using the modified technique is roughly half of that using the conventional technique for both second and third harmonic measurements.

(3) The experimental results obtained here were using surface acoustic waves, but the method to minimize the couplant effect is equally applicable for nonlinearity measurements using longitudinal waves.

TABLE II. Third harmonic nonlinearity measurements at four different input signals.

Input	90 V		125 V		153 V		183 V	
	γ_3^{con}	γ_3^{mod}	γ_3^{con}	γ_3^{mod}	γ_3^{con}	γ_3^{mod}	γ_3^{con}	γ_3^{mod}
Mean	56.18	15.88	51.01	15.40	47.54	16.12	44.13	16.67
Standard deviation	11.65	1.58	9.98	1.56	9.46	1.59	8.84	1.61
COV	0.2073	0.0995	0.1957	0.1014	0.1989	0.0984	0.2003	0.0969

[1] M. Klesnil and P. Lukáš, *Fatigue of Metallic Materials* (Elsevier, New York, 1992).

[2] S. Suresh, *Fatigue of Materials* (Cambridge, New York, 1998).

[3] A. Hikata, B. B. Chick, and C. Elbaum, "Effect of dislocations on finite amplitude ultrasonic waves in aluminum," Appl. Phys. Lett. **3**, 195–197 (1963).

[4] T. Suzuki, A. Hikata, and C. Elbaum, "Anharmonicity due to glide motion of dislocations," J. Appl. Phys. **9**, 2761–2766 (1964).

[5] A. Hikata and C. Elbaum, "Generation of ultrasonic second and third harmonics due to dislocations," Phys. Rev. **144**, 469–477 (1966).

[6] J. H. Cantrell and W. T. Yost, "Acoustic harmonic generation from fatigue-induced dislocation dipoles," Philos. Mag. A **69**(2), 315–326 (1994).

[7] J. H. Cantrell, "Substructural organization, dislocation plasticity and harmonic generation in cyclically stressed wavy slip metals," Proc. R. Soc. London, Ser. A **460**, 757–780 (2004).

[8] S. S. Kulkarni, L. Sun, B. Moran, S. Krishnaswamy, and J. D. Achenbach, "A probabilistic method to predict fatigue crack initiation," Int. J. Fract. **137**, 9–17 (2006).

[9] J. K. Na and J. H. Cantrell, "Linear and nonlinear ultrasonic properties of fatigued 410 Cb stainless steel," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1996), Vol. 15, 1347–1352.

[10] W. T. Yost, J. H. Cantrell, and J. K. Na, "Nonlinear ultrasonic pulsed measurements and applications to metal processing and fatigue," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2001), Vol. 20, pp. 1268–1275.

[11] J. Frouin, S. Sathish, and J. K. Na, "Real-time monitoring of acoustic linear and nonlinear behavior of titanium alloys during low-cycle fatigue and high-cycle fatigue," Proc. SPIE **3993**, 60–67 (2000).

[12] J. H. Cantrell and W. T. Yost, "Nonlinear ultrasonic characterization of fatigue microstructures," Int. J. Fatigue **23**, S487–S490 (2001).

[13] W. L. Morris, O. Buck, and R. V. Inman, "Acoustic harmonic generation due to fatigue damage in high-strength aluminum," J. Appl. Phys. **50**(11), 6737–6741 (1979).

[14] D.J. Barnard, L. J. H. Brasche, D. Raulerson, and A. D. Degtyar, "Monitoring fatigue damage accumulation with Rayleigh wave harmonic generation measurements," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2003), Vol. 22, pp. 1393–1400.

[15] C. Woodward, K. R. White, D. V. Jauregui, and J. Stauffer, "Nonlinear ultrasonic evaluation of concrete microcracking", in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2004), Vol. 23, pp. 1022–1026.

[16] J. Krautkrämer and H. Krautkrämer, *Ultrasonic Testing of Materials* (Springer-Verlag, Berlin, 1990).

[17] J. D. Achenbach, I. N. Komsky, Y. C. Lee, and Y. C. Angel, "Self-calibrating ultrasonic technique for crack depth measurement," J. Nondestruct. Eval. **11**(2), 103–108 (1992).

[18] A. Hikata and C. Elbaum, "Generation of ultrasonic second and third harmonics due to dislocations," Phys. Rev. **144**(2), 469–477 (1966).

[19] J. H. Cantrell, Fundamentals and applications of nonlinear ultrasonic NDE, in *Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization*, edited by T. Kundu (CRC Press, Boca Raton, 2003).

[20] K. E.-A. Van Den Abeele, P. A. Johnson, R. A. Guyer, and K. R. McCall, "On the quasi-analytic treatment of hysteretic nonlinear response in elastic wave propagation," J. Acoust. Soc. Am. **101**, 1885–1898 (1997).

[21] E. L. Adler, E. Bridoux, G. Coussot, and E. Dieulesaint, "Harmonic generation of acoustic surface waves in $Bi_{12}GeO_{20}$ and $LiNbO_3$," IEEE Trans. Sonics Ultrason. **SU-20**(1), 13–16 (1973).

[22] D. C. Hurley, "Nonlinear propagation of narrow-band Rayleigh waves excited by a comb transducer," J. Acoust. Soc. Am. **106**(4), 1782–1788 (1999).

[23] A. P. Mayer, "Surface acoustic waves in nonlinear elastic media," Phys. Rep. **256**, 237–366 (1995).

[24] N. Kalyanasundaram, "Nonlinear surface acoustic waves on an isotropic solid," Int. J. Eng. Sci. **19**, 279–286 (1981).

[25] R. W. Lardner, "Nonlinear surface waves on an elastic solid," Int. J. Eng. Sci. **21**, 1331–1342 (1983).

[26] D. F. Parker, "Waveform evolution for nonlinear surface acoustic waves," Int. J. Eng. Sci. **26**, 59–75 (1988).

[27] E. A. Zabolotskaya, "Nonlinear propagation of plane and circular Rayleigh waves in isotropic solids," J. Acoust. Soc. Am. **91**, 2569–2575 (1992).

[28] H. Ogi, M. Hirao, and K. Minoura, "Noncontact measurement of ultrasonic attenuation during rotating fatigue test of steel," J. Appl. Phys. **81**(8), 3677–3684 (1997).

[29] H. Ogi, M. Hirao, and S. Aoki, "Noncontact monitoring of surface-wave nonlinearity for predicting the remaining life of fatigued steels," J. Appl. Phys. **90**(1), 438–442 (2001).

Efficiency of time-reversed acoustics for nonlinear damage detection in solids

Antonio S. Gliozzi, Michele Griffa, and Marco Scalerandi[a]

CNISM- Dip. Fisica, Politecnico di Torino, Corso Duca degli Abruzzi 24-10129 Torino, Italy

Time-reversed acoustics (TRA) has been developed in the last few years as a powerful tool for several applications, based on the theoretical properties emerging from the time reversal invariance of the wave equation. TRA is expected to be a good basis for the development of imaging techniques in the field of nondestructive evaluation. For this purpose, however, data processing is necessary to discriminate between images due to defects (in general nonlinear scatterers) and images due to linear inclusions, boundaries, etc. We propose here an approach based on the filtering of the time signals. The image of the scatterer is obtained through numerical simulations of the back propagation in a fictitious reference specimen. We validate the approach using the inversion of synthetic data. We also estimate the robustness of the procedure in the presence of constraints that can occur in any experimental procedure. © 2006 Acoustical Society of America.
[DOI: 10.1121/1.2345955]

I. INTRODUCTION

Time-reversed acoustics (TRA) refers to a process in which a propagating elastic wave field (e.g., a vibration velocity field) is reversed in time and irradiated back toward the source from which it was generated.[1,2] Since time reversal invariance[1] is verified by the wave equation, at least in linear, low attenuative, media, the irradiated field is expected to focus back on the source and, generally, reconstruct the signal.[3] The robustness of time reversal is corroborated by the observation that multiple scattering improves the quality of the procedure.[4,5] An additional advantage of TRA is also its potentiality in obtaining super-resolution.[6,7]

Furthermore, the validity of time reversal has been proven in solids.[8–10] More recently, Tanter *et al.*[11] and Cunningham *et al.*[12] have proven reversibility in nonlinear elastic media, at least under certain conditions.

As a consequence of the many theoretical expectations, TRA has been widely applied in different fields such as geophysics,[13–15] medical imaging[16] and applications,[17] oceanography,[18] wave guides,[19] acoustic tomography,[20] etc.

In this context, we are mostly interested in applications of TRA to defect imaging in solids. The idea stems from the consideration that in a TRA experiment the irradiated signal focuses back onto the source of the original signal. Defects in the propagation medium, behaving like scatterers, are expected either to be additional sources in the wave propagation or to partly break time reversal. One major task, however, remains to be solved for achieving optimal application of TRA to defects imaging and detection. It is indeed important to be able to discriminate images produced by real defects, which in general have nonlinear properties, from images due to linear inclusions, local changes in the modulus or boundaries, which also behave as scatterers. The so-called Decomposition de l'Operateur de Retournement Temporal (DORT) method[21] may be implemented to obtain selective focusing on different scatterers. Nevertheless, for early damage detection, it is often necessary to focus on extremely weak scatterers, and this may be particularly difficult. Addressing this issue is the major goal of the present contribution.

Applications of TRA to imaging in the field of nondestructive evaluation have been recently proposed.[22–24] Aluminum plates[25] and composite panels[26] have been studied by comparing the TRA reconstructed signal with the input signal. Any distortion is claimed to be an indicator of the presence of wave scattering due to the presence of damage. TRA has also been applied to detect damage in complex geometrical structures.[27] In particular, simulations have shown that the energy of the reversed wave field focuses on the scatterers, hence indicating a novel imaging procedure.[9,10,28]

In this paper, we suggest a novel procedure to obtain selective focusing on nonlinear scatterers. We also introduce additional novelties with respect to the existing approaches. First of all, we consider systems in which the surfaces of the specimens, or part of them, are not easily accessible. As a consequence we will not exploit, as is usually done, all the information of the received signals along a given direction, which may be one side of a plate[23] or a radial circumference for cylindrical specimens.[27]

In addition, we consider an experimental setup in which effects of boundaries of the specimen are important. As a consequence, we use information intrinsic also to the transmitted signals and not only to the backscattered ones. Finally, we underline that the geometry used here does not have any partial/complete symmetry. For instance, the time of flights from the defect to the receivers differs largely from one receiver to another, which may render the choice of the reversal time window a crucial point in the procedure. This problem does not appear when receivers, as usually done, are

[a]Author to whom correspondence should be addressed. Electronic mail: marco.scalerandi@infm.polito.it

arranged with a certain symmetry with respect to the scatterer (e.g., scatterer in the center of the specimen and receivers on the surfaces).

We make use of computational time reversal as a technique for imaging: After filtering of signals obtained from forward propagation in a damaged specimen, the results of simulations of the back propagation lead to the detection and characterization of regions where scatterers are located. The approach is tested by using synthetic data. In the next section, we describe the considered specimen, which includes both a linear inclusion and a nonlinear scatterer, and the nonlinear hysteretic model used to produce synthetic data.[29,30] We remark that the procedure and validity of the results proposed here are independent from the choice of the nonlinear elastic model. Forward data are indeed expected to be provided by a real experiment. In Sec. III, we present different procedures for treating signals before reversal.[31] In Sec. IV, we evaluate the robustness of the procedure by introducing some of the possible intrinsic sources of constraints/inaccuracies in real experiments, such as noise in the signals or weak sources of nonlinearity distributed in the specimen. Finally, in Sec. V, we further validate the proposed approach with a full three-dimensional (3D) simulation, with a configuration closer to that of a real experiment in which, e.g., only one surface of the specimen is available for transducers placement.

II. FORWARD PROPAGATION

A time-reversed acoustics (TRA) experiment consists of three successive steps:

- Generation of acoustic/ultrasonic waves in a specimen and detection of the signal (e.g., vibration velocity) as a function of time in several locations by means of a large number N of transducers acting as receivers and possibly embedded in the system. If that is not possible, a scanning laser can supply the signals from N different positions. Let us call $\boldsymbol{u}_j(t)$ the signal received by the transducer j $(j=1,\ldots,N)$ located in the position (x_j, y_j); note that the signal is a vector.
- Processing and time reversal of the received signal at each receiver. Let us call $\boldsymbol{v}_j(t)$ the processed signal.
- Reinjection of the time reversed signals in the specimen by each transducer, now used as a transmitter.

As discussed in the next section, the latter two points are numerically performed, while the former is expected to be experimentally realized. In this sense, we use numerical simulations as tools for post-processing experimental data and define the imaging procedure. Simulations are indeed the only tool for implementing successfully the procedure proposed here when nonlinear scatterers are located in the bulk of the specimen. On the contrary, when the focusing point is localized on the surface, the third step can be realized experimentally, by scanning the surface of the specimen, e.g., with a laser interferometer, to detect the focal point.

Here we produce synthetically the set of received signals, using a 2D model, which corresponds to an approxima-
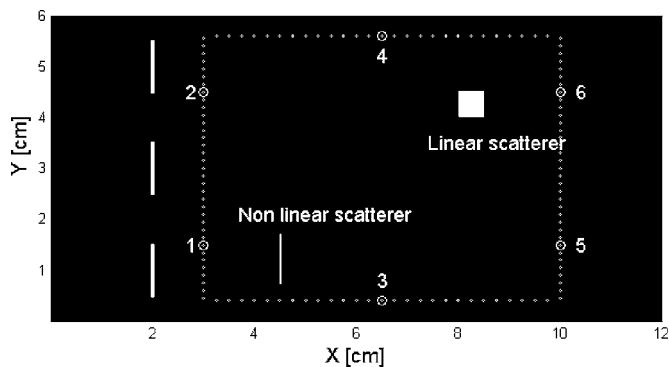
FIG. 1. Representation of the plate used in the simulations. Transducers are indicated with a vertical line along their extension. Receivers are represented by small dots and few of them are marked with a number.

tion of a 3D specimen with no out-of-plane components of either stresses or strains. A few results using a 3D specimen will also be provided in Sec. V.

A. The specimen

For our simulations a 2D aluminium plate has been chosen as a test specimen (see Fig. 1). It has dimensions 12 cm × 6 cm and elastic properties given by the Lamè constants $\lambda = 51.08$ GPa and $\mu = 26.32$ GPa. Density is assumed to be $\rho = 2700$ Kg/m^3. Consequently the wave velocity in the material is $v_L = 6176$ m/sec and $v_S = 3122$ m/sec for the longitudinal and shear components, respectively. Attenuation in the material is chosen to be very high (to have the worst condition for time reversal) and therefore only few reflections from the boundaries will be considered. The boundaries of the specimen are considered fixed and not accessible to receivers, to simulate a specimen embedded in a larger structure.

Two scatterers are located as shown in Fig. 1. The larger one is a square-shaped linear inclusion of size 5 mm × 5 mm, with the same density as the bulk and a 20% reduction in both Lamé constants. The defect is a small nonlinear scatterer with length 10 mm and width 0.1 mm. The two have been located in arbitrary positions, far from each other. The configuration with the scatterers far from the center has been chosen in order to be far from the optimal configuration for reconstruction (e.g., defects in the center and/or equidistant from receivers). In fact, here influence from the boundaries is present in the signals.

The specimen is equipped with three extended transducers (size 10 mm × 1 mm). Each of them is designed to force a stress into the specimen with Gaussian modulated profile (with center angular frequency $\omega = 1.2$ MHz). The choice of three transducers is motivated by the need to have a uniform coverage of the specimen by the wave field. Results of the TRA procedure however do not seem to be affected by the number/location of the transducers and by the signal type.

Finally, $N = 124$ receivers are placed in a square arrangement around the zone that have to be investigated. A few receivers have been labeled with a number of reference. As will be discussed later, not all the received signals will be used for temporal inversion.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Gliozzi *et al.*: TRA for nonlinear damage detection   2507
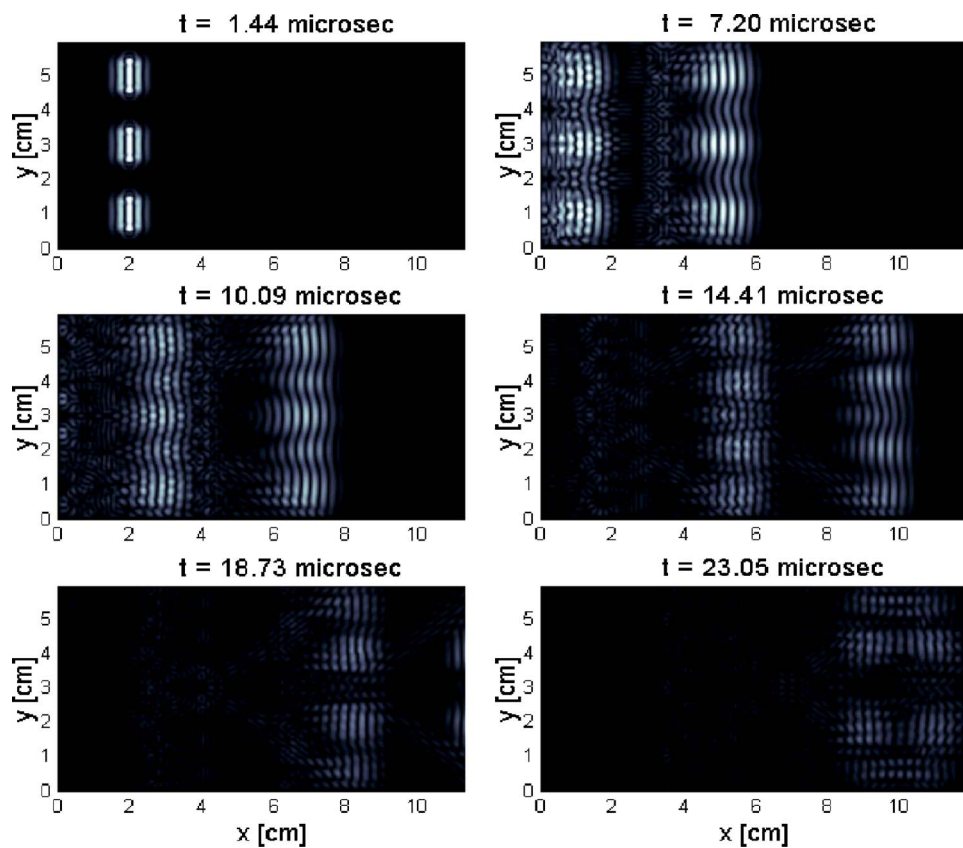
FIG. 2. Snapshots of the vibration velocity at successive times during the forward propagation in the plate with a linear and a nonlinear scatterer.

B. The model

To simulate the propagation of ultrasonic waves in the specimen of Fig. 1, a two-dimensional (2D) approach to the simulation of ultrasonic wave propagation in nonclassical nonlinear (NCNL) media has been used.[32] The method is described elsewhere[33] and its correctness demonstrated by qualitative and quantitative comparison with experimental data.[34–37] In particular, the model predicts mode conversion between shear and longitudinal waves, both due to the propagation of a nonplane wave[34] and both that generated at the reflection from an interface between different materials.[36] The issue is not discussed further here due to space limitations, but has been widely demonstrated in the mentioned references.

The approach represents the extension to 2D of a previously proposed 1D model,[29,30,38] based on a spring model,[35] applied together with a Preisach-Mayergoyz (PM) space treatment.[39] The material is discretized into a large number of elements (1200×600 in our simulations). Each of these is described by the usual linear elastodynamic equation of state in which stress is proportional to strain through the elastic stiffness tensor. The nonlinearity is introduced in the approach by including nontrivial boundary conditions on each element. This is obtained by defining thin space regions ("interstices") between adjacent cells that may be used to describe imperfections in the material.[35] The choice of a rigid interstice corresponds to imposing continuity of stresses and displacements between cells, i.e., absence of defects. Elastic interstices describe linear inclusions or delaminations.[40] Interstices switching from a rigid to an elastic state and vice versa during the crossing of the propagating wave describe a

nonlinear inclusion. Finally, when a PM space is introduced in order to model the transitions between the two states, regions with nonlinear hysteretic properties are modeled.[29,30,33]

C. Results

The chosen configuration of three transducers produces the propagation of a wave front (close to a plane wave) in the specimen, so that almost the whole of the specimen is excited with the same stress amplitude. In Fig. 2, snapshots of the vibration velocity at selected times during the propagation are reported. As expected, both the linear and the nonlinear inclusions are too small to produce appreciable effects.

The signals at two selected receivers (R) are reported in Fig. 3 and compared with the signals obtained in the absence of any inclusion/defect (reference specimen). The arrivals of the waves after multiple scattering from the specimen boundaries are clearly visible. For instance, the signal at 10 μsec for $R1$ corresponds to the wave reflected by the left edge of the specimen; the signal at 22 μsec at $R6$ corresponds to the superposition of the wave fronts reflected from the left and right boundaries. The correspondence with the reference signal is very good, confirming the difficulty to detect the presence of the two inclusions from the forward propagation only. A slight mismatch is only visible at $R6$, which is located very close to the linear scatterer. We deduce the greater influence of the inclusion on the propagation compared to the influence of the defect.
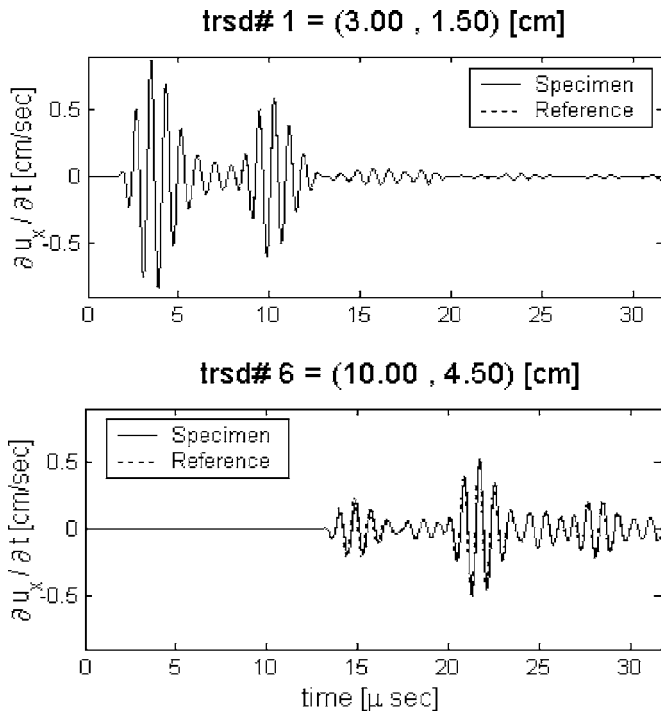
FIG. 3. Temporal signals recorded at the receivers number 1 (R1) and 6 (R6), see Fig. 1. The horizontal component of the vibration velocity vector is reported. The reference data refer to results for a homogeneous specimen, i.e., without the two scatterers.

III. IMPLEMENTATION OF TRA

To implement computational time reversal, the signals $u_j(t)$ recorded at each transducer j should be processed, time reversed, and reinjected into a reference fictitious specimen with the same geometry and boundary conditions of the previous one, but without inhomogeneities and defects. Injection is applied at the same time from all receivers, now used as actuators. The phase delay between the reinjected signals $v_j(t)$ should produce focusing on the position of the source of the original wave. It is to be noted that, in our problem, we can consider different sources: The three emitters used in the forward propagation problem, the linear, and the nonlinear scatterers and the boundaries.

A. Signals processing

Signals generated in the forward propagation have to be properly treated before being reinjected in the specimen. The very first step consists in the selection of the time window, characterized by a length Δ of the time signal to be reversed and an initial time T. Here we use a time window of 18 μsec, starting from 7 μsec. It is worth noting that the choice of the time window is not a trivial issue. More sophisticated choices could be made accounting for the fact that, due to the different distances of the receivers from the scatterers/sources, the optimal selected time window can in principle be different for each receiver. The issue is, however, beyond the scope of this contribution.

As a second step, the selected portion of the signal should be time reversed and additional processing may be introduced, in order to exploit the properties of time reversed propagation in detecting the positions of scatterers in a se-

lective way, distinguishing between boundaries, linear inclusion, and nonlinear defect. We propose three different treatments of the received signals, denoted in the following as *reversal*, *subtraction*, and *filtering* processes. The procedures have been designed so that in each one the portion of the signal due to the larger source, already localized in the previous processing step, is eliminated. We define an operator $\check{R}$ as $\check{R}(u_j)(t) = u_j(T + \Delta - t)$. The three treatments proceed as follows:

- Reversal processing: It corresponds to the usual TRA procedure. The injected signal is defined as $v_j(t) = \check{R}(u_j)(t)$, $0 < t < \Delta$. Unfortunately, as visible from Fig. 3, the three transducers are by far the stronger sources producing the signal $u_j(t)$. Therefore, in this standard procedure the portion of the signal due to scatterers is negligible.

- Subtraction processing: Before time reversal, we subtract from the received signal the reference signal $u_{jo}(t)$, i.e., the signal recorded at the same positions in the absence of any scatterer. It follows: $v_j(t) = \check{R}(u_j)(t) - \check{R}(u_{jo})(t)$. This processing analysis aims to eliminate the portion of the signal generated from the transducers only. Unfortunately, such an approach requires a baseline reference, which is not always available with the desired accuracy. It follows that the main advantage of using TRA for imaging purposes is lost. Also, in this procedure most of the signal is expected to be due to the linear scatterer, rather than to the (more interesting) nonlinear one.

- Filtering processing: The final goal of processing is to eliminate the portion of the signal due to the linear scatterer. For this purpose, considering that the nonlinear defect is the only source of nonlinearity in the specimen, we expect it to be the source for the portion of the signal containing higher order harmonics. Therefore, before reversal, we process the signal using a filter, which allows the elimination of all the components close to the fundamental angular frequency ω (center frequency of the injected signal). It follows that, without any need of a reference signal, only components generated by the defect are maintained in the received signal. The injected signal has the form: $v_j(t) = \check{R}(u_j^F)(t)$, where F denotes that the signal has been filtered. In the following a bandpass filter between 1.8ω and 4ω has been used.

Results for the three procedures are reported in Fig. 4 for selected receivers. The processed signals are similar when only a reversal is performed, reflecting the symmetry in the placement of the three transducers. On the contrary they are very different in the other two cases (bottom two rows), reflecting the asymmetries in the placement of the scatterers. Indeed, the signal at $R5$, which is aligned with the nonlinear scatterer (see Fig. 1), is almost null when the subtraction processing is applied, while rather large when filtering. The opposite occurs for $R6$, aligned with the linear inclusion. Note in Fig. 4 the different scales on the y axes. While after
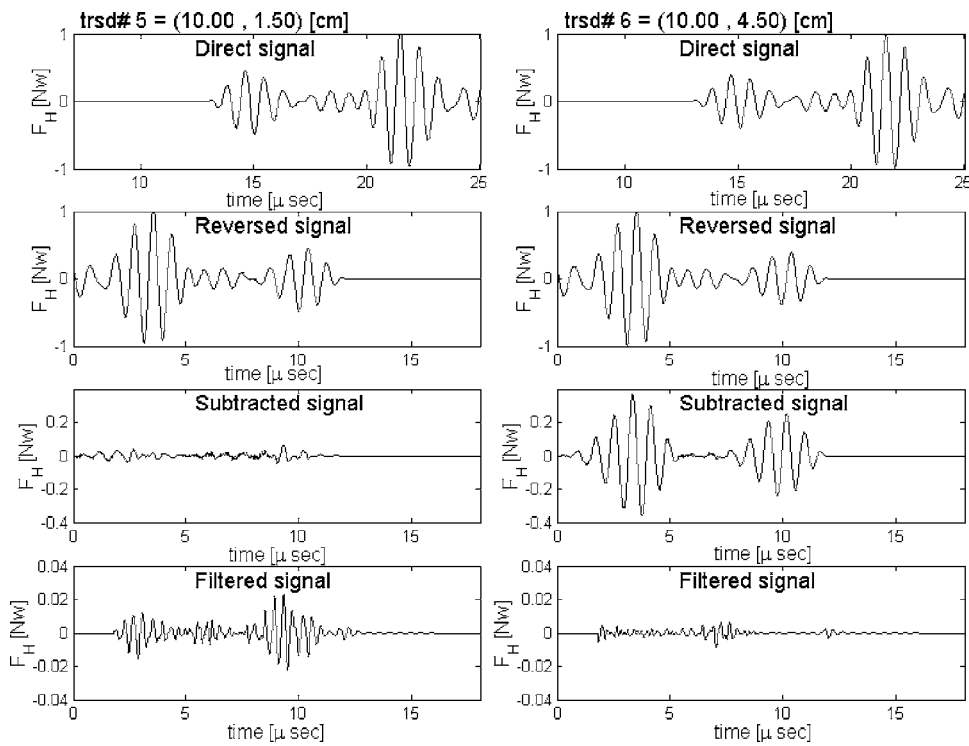
FIG. 4. Results of the processing of the signals at receivers number 5 and 6 (see Fig. 1) before reinjection into the specimen. The horizontal component of the force per unit surface is reported in the selected time window. (a) First row: received signal; (b) second row: reversal of the received signal; (c) third row: reversal of the signal obtained subtracting the reference signal; and (d) fourth row: reversal of the band-pass filtered received signal.

subtraction about 25% of the signal is used for reversal, only about 2% remains after filtering. It follows that noise may play a crucial role, as discussed in Sec. IV.

B. Results

As a final step of the TRA implementation, the processed signals $v_j(t)$ are injected back into the specimen at the same time from all transducers. Simulations of the propagation of the time-reversed field are performed using a homogeneous linear specimen (no scatterer), in order not to introduce any bias from an *a priori* knowledge of the defect location. Note that here the vectorial form of the reversed signal is maintained in the injection (i.e., both components of the received signals are reinjected). Receivers in experiments may not be able to keep the vectorial information. However, as discussed later, the imaging procedure is still effective.

In Fig. 5, we plot the map of the velocity field at the time corresponding to best focusing during the reversed propagation for the three processing procedures described in the previous subsection: reversal, subtraction, and filtering, from top to bottom. As expected, albeit at different times, best focusing is obtained on the three transducers, on the linear inclusion and on the nonlinear scatterer, respectively. In each of them, the temporal evolution of the wave field does not indicate evidence of focusing on other sources except the ones indicated in the plot. A very slight evidence of focusing on the nonlinear inclusion is only present in the wave field corresponding to the subtracted reversal process, which indeed was to be expected. The reduced focusing of energy on the central transducer in the upper plot is due to the choice of the distribution of receivers.

In this contribution, we are interested in the last proposed procedure only, which is novel, to our knowledge, with respect to existing TRA approaches and does not re-

quire any reference signal. Therefore, we further discuss,
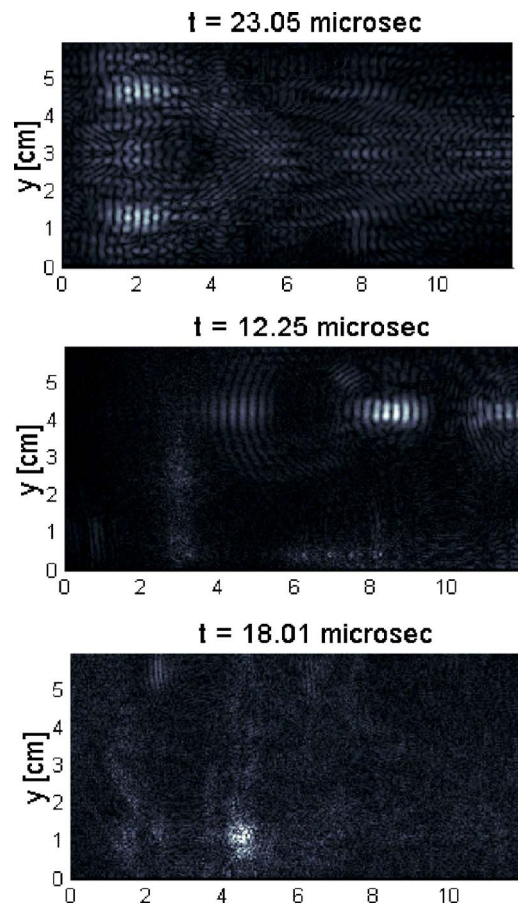
FIG. 5. Snapshots of the vibration velocity as obtained in the simulations of the reversed signals at the time corresponding to best focusing for the three implementations proposed. First row: reversal; second row: subtraction; third row: filtering.
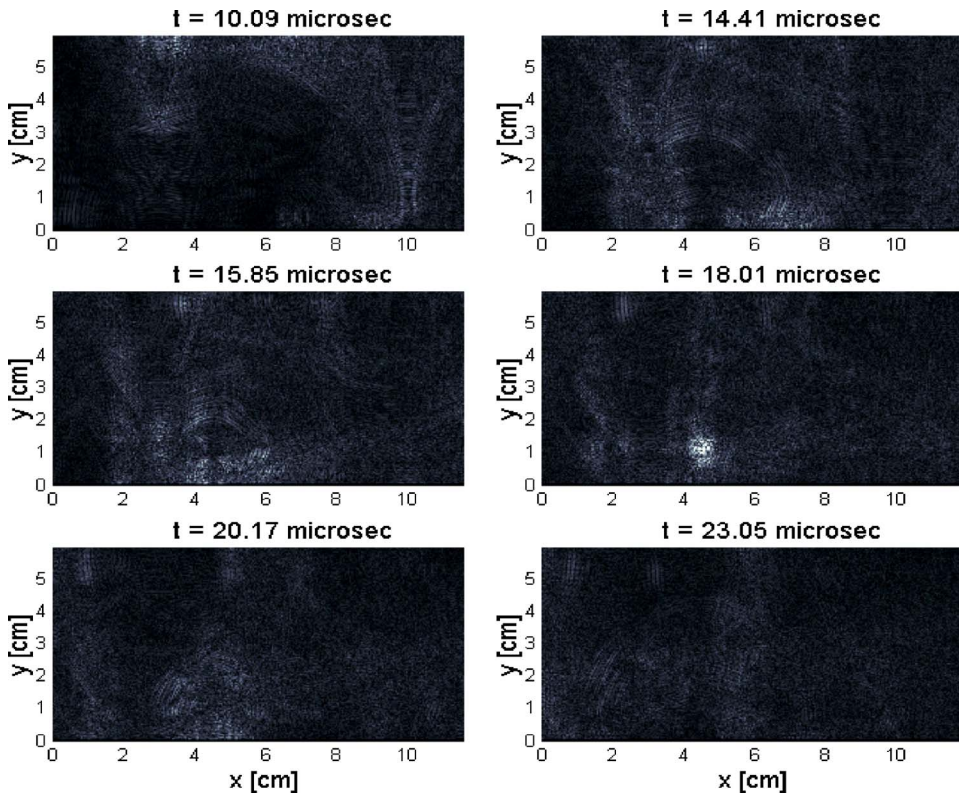
FIG. 6. Snapshots at successive times of the velocity wave field during back propagation in the case of filtering.

here and in the following, the case of filtering. In Fig. 6 snapshots of the velocity field are reported at selected times during reversed propagation. The time-reversed wave field appears at positions farther from the source (the nonlinear scatterer) first, i.e., the upper and right rows of receivers. At a later time (14–15 μsec), the receivers close to the defect start playing a role, but the wave field remains unfocused. Progressive focusing occurs between 16 and 18 μsec revealing a reasonably good image of the scatterer. At later times, defocusing occurs again. No evidence of focusing in regions far from the scatterer (i.e., phantom images) is present.

C. Improvements of the imaging procedure

Images of the wave field at the time corresponding to best focusing (see Fig. 5) constitute a first indicator of the quality of the imaging procedure. In these plots, however, several pieces of information are lost. For example, the possible focusing on other scatterers and the images of the ray paths from the receivers to the focusing point. To capture these details an alternative mapping of the wave field data can be proposed as follows.

Let us call $w(x,y;t)$ the function describing one variable (vibration velocity, strain, stress, etc.) in any position (x,y) and at any time (t). We introduce the following function:

$$M(x,y) = \max[w(x,y;t)]_t \qquad (1)$$

which represents, in any position, the maximum value the variable w assumes in time. The function M is expected to present maxima wherever focusing occurs and also secondary maxima where the unfocused paths from the N receivers interfere coherently toward the focusing point, hence identifying the ray paths. Here, we have chosen the variable w to

be the modulus of the local displacement vector field.

An image of the map M corresponding to the same simulation of Fig. 6 is reported in the upper row of Fig. 7. The map clearly indicates focusing on the nonlinear scatterer and a maximum in the position of the receivers which are mostly contributing to the reconstruction (i.e., receivers close to the scatterer). Some ray paths are also identified, although not particularly meaningful here.

To improve the imaging procedure, we note that Fig. 7 shows that only few transducers contribute to the reconstruction of the image. Also, the time series in Fig. 4 illustrate that some transducers reinject a signal of very small amplitude (e.g., $R6$). Since focusing is determined mostly by keeping proper phase delays between signals injected from different positions, independent of amplitudes, we introduce the possibility of amplifying the injected signals differently at different positions. Therefore, we introduce a further signal processing step: $z_j(t) = A_j * v_j(t)$, where A_j is the amplification factor. To obtain equal contributions from all transducers, we define $A_j = 1/\max[\|v_j(t)\|]_t$. The corresponding z_j signals are therefore each normalized to one.

Results of the simulation using normalized signals are reported in the second row of Fig. 7, where the map of the function M is reported. The improvement of the reconstructed image is evident, with a much larger energy being focused on the scatterer than that being injected at the transducers. Ray paths are more meaningful and give indication of the progressive focusing. In the third row of Fig. 7 we also analyze the effect of reinjecting normalized signals using only one component of the recorded wave field (in particular the x component). The reconstructed map remains reasonably good, albeit with a deterioration of the image with respect to the previous case.

Not normalised

Normalised

Only x-component

Averaged
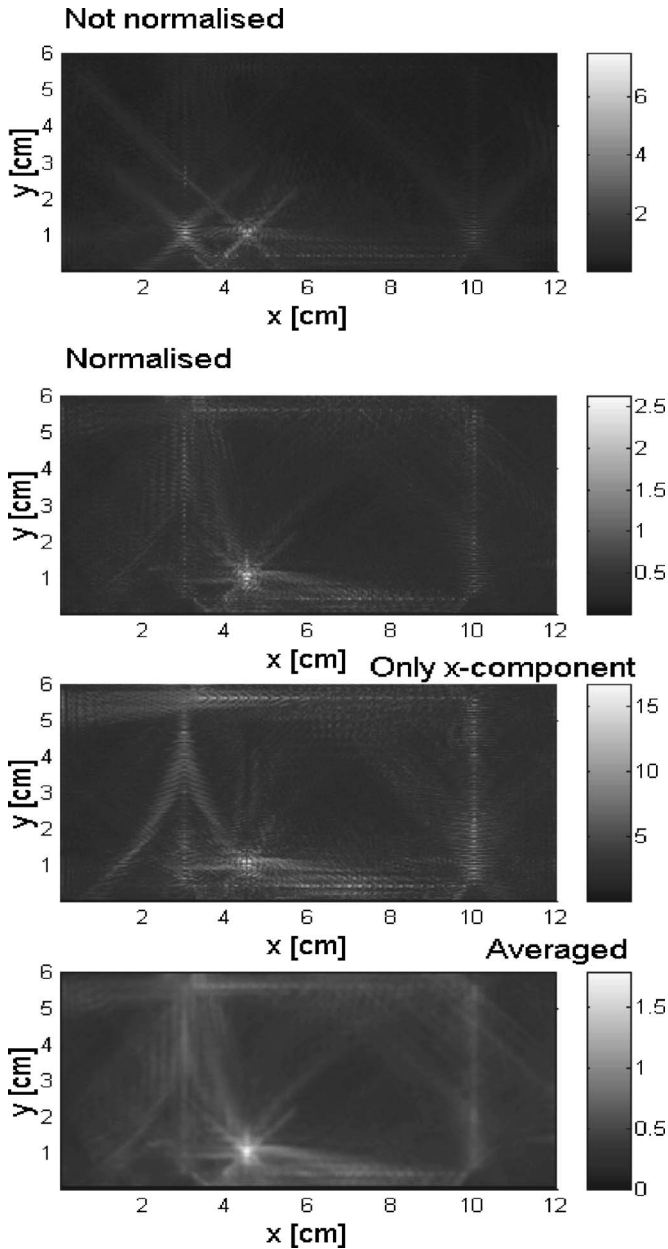
FIG. 7. Maps of the maximum of the displacement [see Eq. (1)] for different implementations of the procedure. (a) First row: standard procedure; (b) second row: normalizing all the injected signals to 1; (c) third row: normalizing the signals and reinjecting only the horizontal component of the recorded signal; and (d) fourth row: averaging the map over a square with size equal to the distance between adjacent receivers.

Finally, in the bottom row of Fig. 7 we also show how a mathematical smoothing on the map M leads to a significant improvement of the quality of the image of the defect location. This plot corresponds to the same image of the second row, but averaged on a square of size corresponding to the distance between two adjacent receivers (3 mm).

IV. ROBUSTNESS OF THE IMAGING PROCEDURE

This section is devoted to testing the robustness of the filtering procedure for the purpose of nonlinear defect imaging. We investigate the effects on the quality of the image of various constraints/errors that may be intrinsic to the experi-

mental procedure of detection of the forward signals. In all cases, we show the robustness of the imaging process to be valid only up to a certain point.

All results reported in this section refer to plots of the matrix M [see Eq. (1)], obtained implementing the filtering procedure, with normalized signals and maintaining the vectorial information in the backward injection.

A. Effects of the number of transducers

In Fig. 8, we report the reconstructed images obtained using a different number of transducers. Except for the plot in the bottom-right corner, transducers are always taken equidistant on the four sides of a square. In the bottom-right plot, only transducers localized on the two vertical arrays are used, which may be an easier configuration from an experimental point of view.

As expected the quality of the image decreases with a decrease in the number N of transducers available. In fact, decreasing N, the maxima of energy become more and more focused on the transducers positions: Fewer transducers mean fewer rays converging on the defect location and, consequently, less constructive interference. Nevertheless partial focusing is still evident in the case $N=22$.

Note that the need of a quite large number of transducers should not be considered a technical obstacle. In fact, receivers for the forward propagation may be easily substituted by a scanning laser interferometer, as mentioned. Emitters in the reversed propagation are used only in the numerical simulation, where an arbitrarily large number of emitters can be used without difficulties.

B. Effects of noise in the received signal

Clearly, synthetic data are not affected by noise. On the contrary, in every experimental system, noise is present and contaminates the received signals. To simulate a realistic condition, we artificially perturb the received signals with a randomly generated additive noise before processing:

$$\boldsymbol{u}_j^n(t) = \boldsymbol{u}_j(t) + \gamma_j(t),$$

where $\gamma_j(t)$ is a random white noise signal with zero average. Therefore, for each of its components k, $\langle \gamma_k(t) \rangle = 0$ and $\langle \gamma_k(t') \gamma_k(t) \rangle = (\gamma_0^2/4) * \max[u_k(t)]_t * \delta(t,t')$. Typical effects on the signal are reported in Fig. 9 for the receiver number 1 and two choices of γ_0. As expected, the effect on the received signal (upper row) is small. On the contrary, it increases enormously in the filtered signal (second row). In the case of a larger γ_0, the effect of the noise is comparable with the original signal (see third row).

Despite the large effect of noise, the reconstructed images of the defect remain acceptable up to a noise level of $\gamma_0 = 0.1$ (see Fig. 10). Not surprisingly, the procedure does not work any longer at the largest value of γ_0, where no evidence of focusing is observable. Rather unexpected is the robustness of the procedure at $\gamma_0 = 0.1$, considering the quality of the injected signals. The result is an indicator that the noise, albeit very large, produces an unfocused wave field superimposed to the wave field generated by the nonlinear
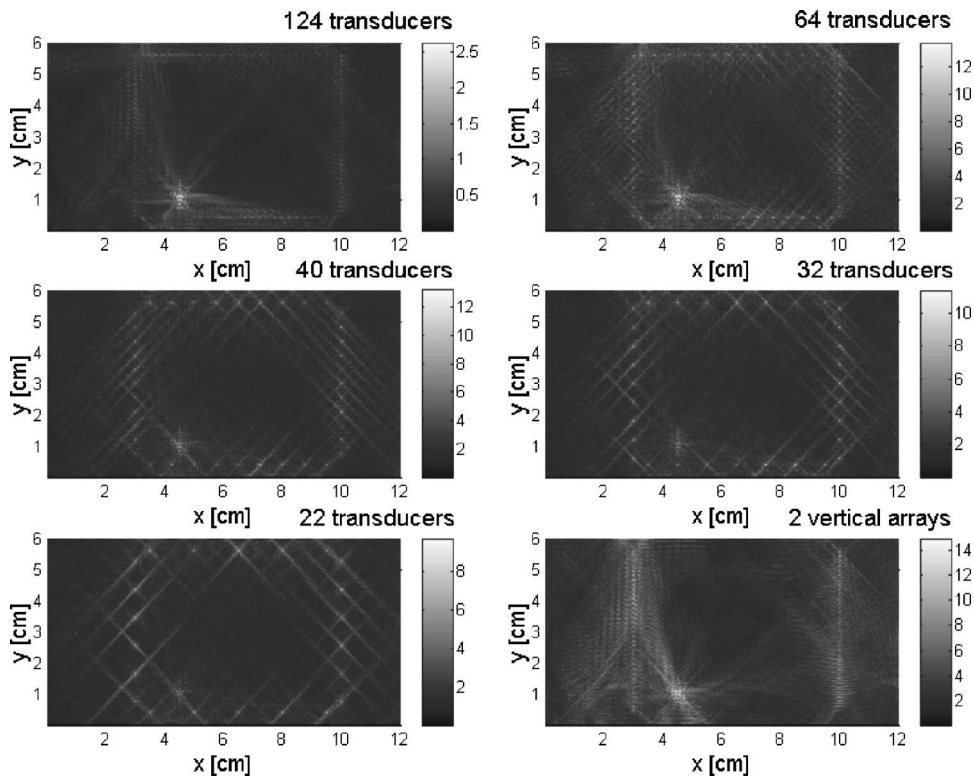
FIG. 8. Maps of the maximum of the displacement [see Eq. (1)] for different numbers of used receivers. The bottom right corner corresponds to the results obtained when only two vertical arrays of receivers are used.

scatterer. The former is distributed on the full specimen, resulting in a local energy which is very small compared to that which focuses on the scatterer.

C. Effects of a distributed weak nonlinearity

Similar encouraging results concerning the robustness of the proposed procedure are also obtained when analyzing the capability of the approach to detect the scatterer in the presence of a distributed nonlinearity. For this purpose we consider a specimen with several weak nonlinear scatterers (about 5–10% of the nonlinearity of the defect) distributed randomly in space. To quantify such distributed nonlinearity, we define a parameter α which defines (as a percent area) the zone in which nonlinear scatterers are distributed, hence giving the density of nonlinear scatterers.

The signals obtained from the simulation of the forward propagation differ only very slightly from the ones obtained in the absence of weak scatterers. On the contrary, such a nonlinearity has an enormous influence on the filtered and reversed signals already at $\alpha=5\%$ (see Fig. 11).

Despite the huge effect on the injected signals, reconstruction of the defect image remains good up to $\alpha=10\%$
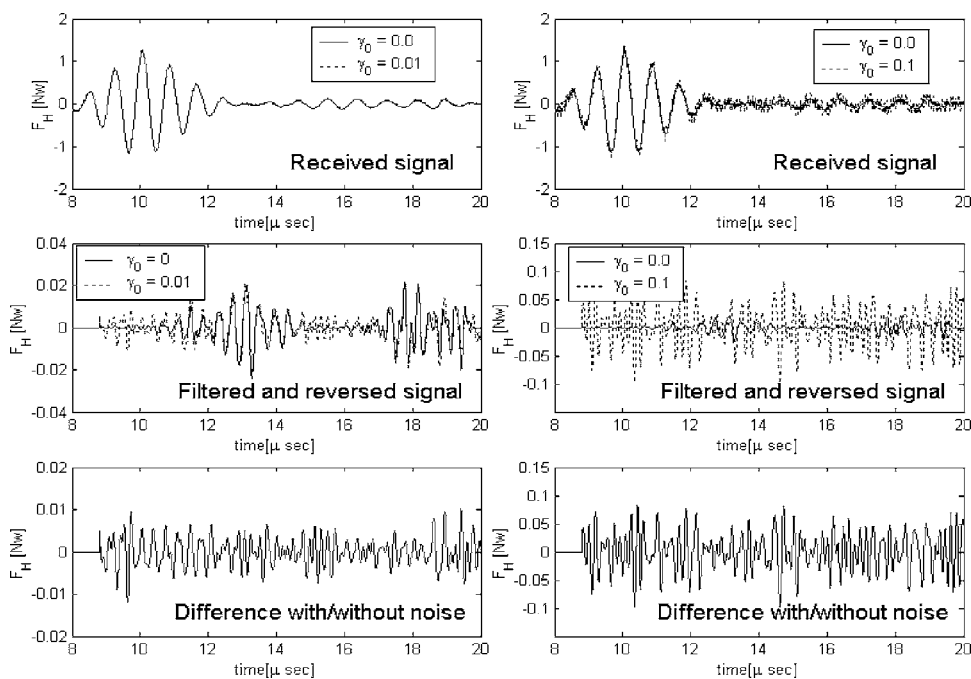
FIG. 9. Comparison of the temporal signals at receiver 1 (see Fig. 1) in the presence or absence of an additive noise on the received signal. First column: weak noise level; second column: large noise level.
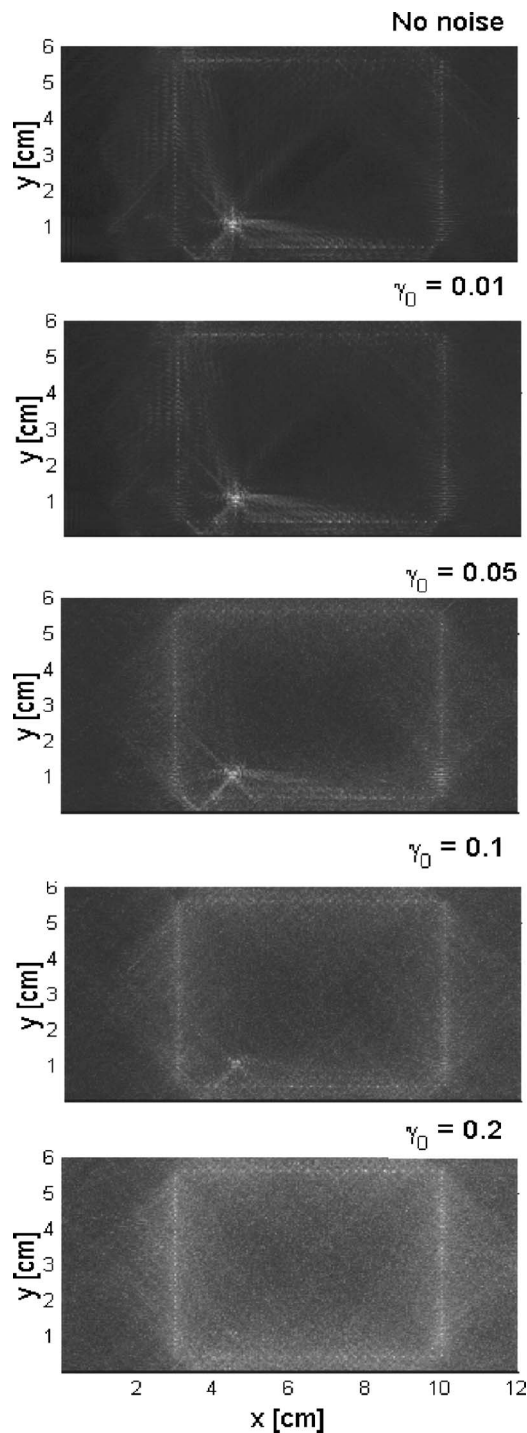
FIG. 10. Maps of the maximum of the displacement [see Eq. (1)] for different noise levels.

FIG. 11. Temporal signals at receiver 1 [see Fig. 1] for different levels of distributed nonlinearity in the specimen.

(see Fig. 12). Note the significant energy focusing close to the transducers used in the forward problem. Such focusing is due to the fact that, in the forward propagation problem, nonlinear scatterers close to the source have been excited at a larger amplitude than scatterers far from it.

V. EXPERIMENTAL FEASIBILITY

So far, we have used a 2D model of elastic (ultrasonic) wave propagation in heterogeneous specimens, which is an approximation of a 3D model of propagation in a very thin plate, under the special conditions of plane stress propagation. Such approximation has been widely used in the literature (see, e.g., Refs. 25–28 and 41–44) to prove feasibility and properties of imaging methods based upon numerical simulation and processing of real data.

Nevertheless, it may be argued if the method proposed here still works in conditions closer to the experimental ones, e.g., when measurements of in-plane vibrations are not possible. Therefore, in this section we present an application of the proposed procedure to a 3D specimen, using the extension of the 2D approach to account also for displacements/stresses in the Z direction.

We consider a thin plate (10 cm × 2.5 cm × 1 cm) with Lamè constants $\lambda = 106$ GPa and $\mu = 71$ GPa. Density is assumed to be $\rho = 8000$ Kg/m^3. A small nonlinear scatterer (5 mm × 3 mm × 2 mm) is located in $x = 5$ cm, $y = 1.5$ cm, and $z = 0.5$ cm. One transducer is localized on the upper surface in $x = 4$ cm, $y = 1.2$ cm, injecting a Gaussian modulated wave with center frequency $\omega = 0.7$ MHz. Forty-five receivers are localized on the upper surface ($z = 0$) along five parallel arrays (see Fig. 13 where the position of the arrays is clear).

The data processing and the simulation of the time reversed propagation is performed following the same procedure as before (filtering with normalization) and a 3D matrix M is defined [analogous to Eq. (1)]. In Fig. 13, slices of the matrix M at different depths are reported (darker grey tones denote larger intensities). On the upper surface ($z = 0$), the position of the receivers is evident. At the same time, we can observe focusing of the energy on the position occupied by
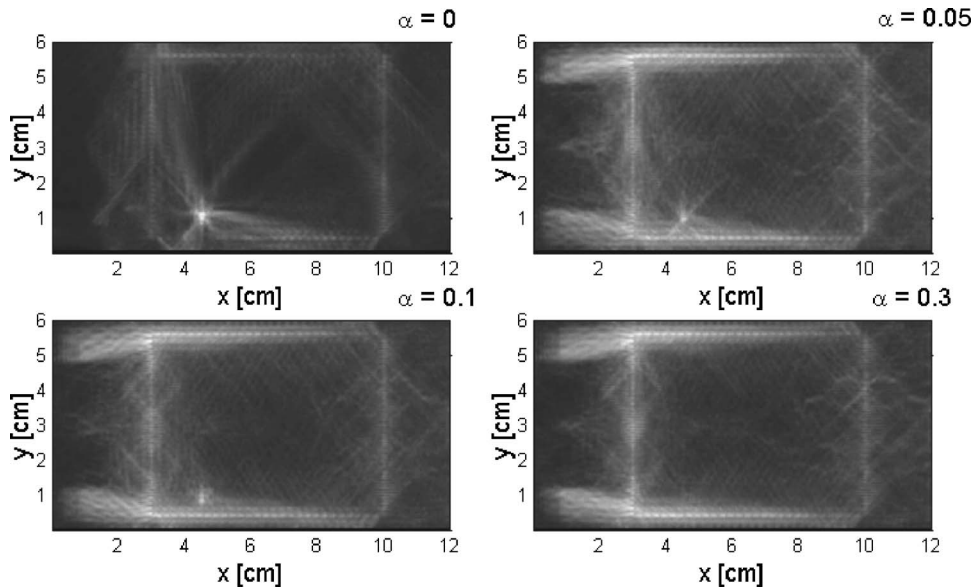
FIG. 12. Maps of the maximum of the displacement [see Eq. 1] for different distributed nonlinearity levels.

the nonlinear scatterer, while no focusing is evident in planes further from $z=5$ mm. Results are in good agreement with that obtained in the 2D approximation (see, e.g., Fig. 7).

A 3D representation of the matrix M is also reported in Fig. 14: We report sections along three lateral surfaces and a diagonal cut from a line of receivers to the defect center. Again, focusing on the scatterer location is visible.

VI. DISCUSSION

All studies reported in the literature point out that TRA, compared with other damage localization methods, remove

the requirements of a reference system for comparison.[24] This is one of its major advantages. Nevertheless, in our opinion, this is not always the case. In fact, increasing the complexity of the specimen, small defects may not be detected in the reconstructed TRA image, due to the presence of stronger scatterers in the sample. Indeed, in this contribution, we have considered the case of weak scatterers, which give an almost negligible signature on the received signals (see Fig. 3). As a result, as shown in Fig. 5, a linear inclusion, or even the presence of boundaries or of the transducers
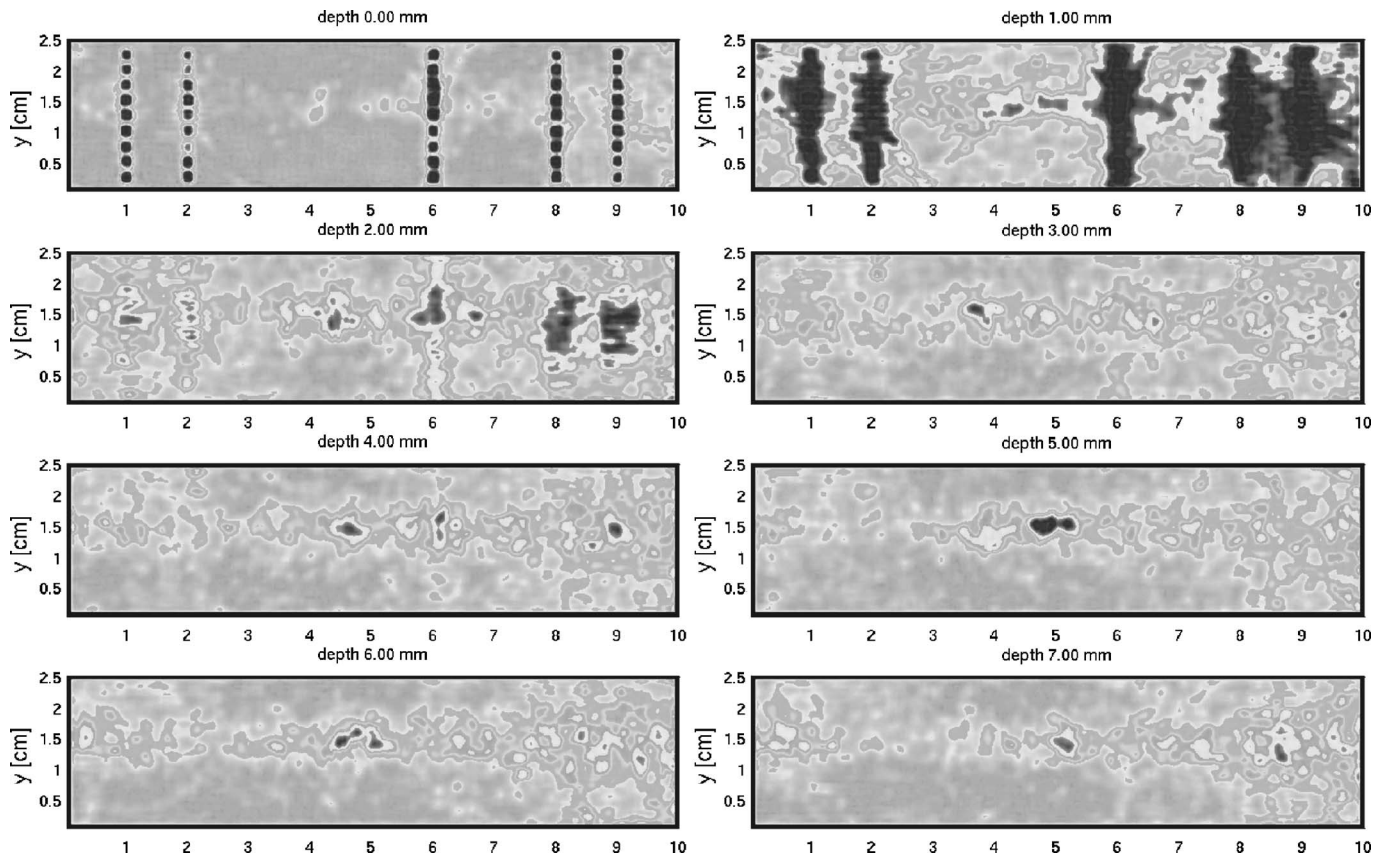
FIG. 13. Maps of the maximum of the displacement in a 3D numerical experiment. The various rows represents cuts at different depths as indicated. The nonlinear scatterer was located in $x=5$ cm, $y=1.5$ cm, and $z=0.5$ cm. Darker grey tones denote larger intensities.
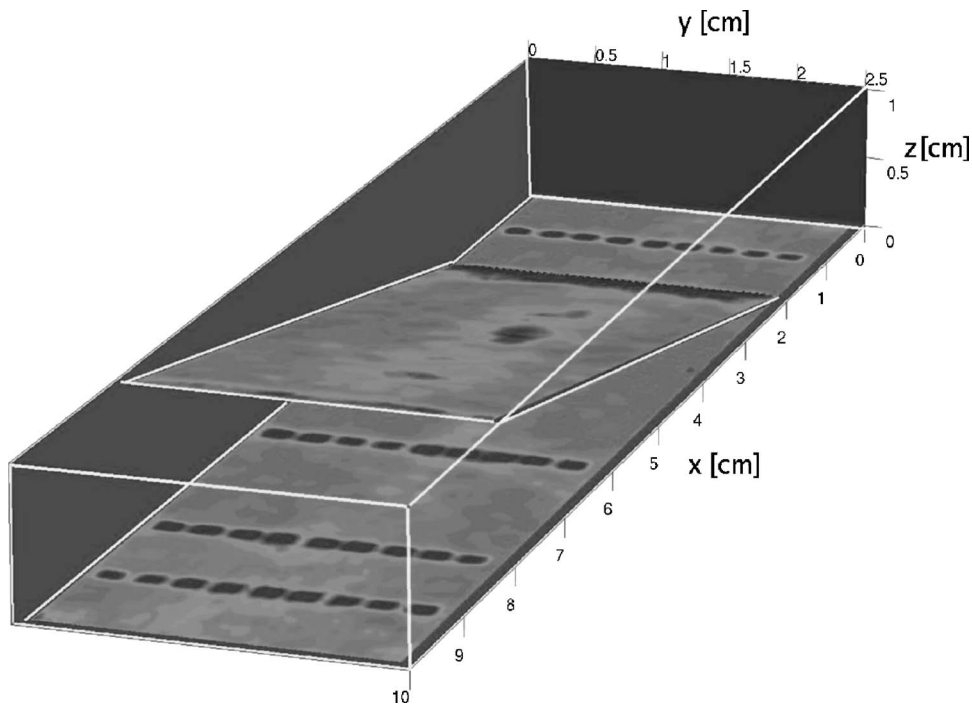
FIG. 14. 3D maps of the maximum of the displacement in the same case as Fig. 13. Four reference planes are reported, the diagonal one intersecting one array of transducers and crossing the center of the nonlinear scatterer. Darker grey tones denote larger intensities.

themselves, may reduce the efficiency of the procedure and make difficult, for traditional TRA, to detect a nonlinear defect.

As a consequence, we have proposed here to exploit features intrinsic in the nonlinear properties of small defects to maintain the possibility to avoid using a reference system, as proposed elsewhere in a different context.[45,46] Indeed, the harmonics content of the received signals may be due only to the presence of nonlinear features in the specimen. It follows that, selecting such higher order components corresponds to subtracting a reference signal, without the need of it. As we have shown, the procedure can be easily implemented and leads to selective focusing on defects with nonlinear characteristics.

The method proposed here has been tested using synthetic data, produced by a numerical code based on a model proven to be reliable in the description of the propagation of elastic waves in nonlinear hysteretic media.[29] In the paper, a 2D configuration has been considered, which is far from being technically realisable (e.g., a laser interferometer can measure the out of plane components of the vibration velocity, fixed boundary conditions are usually not realistic, etc.). Nevertheless, we believe that the theoretical analysis reported here might be an important preliminary step toward a practical realisation and the results obtained for a simple 3D case (Sec. V) further confirm our belief. Furthermore, the capability to suggest novel experimental approaches is, in our opinion, one of the purposes of numerical modeling, as demonstrated by the large number of papers in which new signal processing analysis methods and theories have been tested also using numerical simulations.[23,41–44]

We remark again that imaging, in this contribution, is considered as the postprocessing of the data, realized through the simulation of the backward propagation of the time-reversed signals. To implement such an approach, is therefore necessary to record a sufficient number of forward signals from the experiment and to provide a detailed information about the elastic and geometrical properties of the specimen (in its intact state), in addition to the details about the boundary conditions adopted in the experiment.

Several issues have also been raised in the paper and probably need further attention when lab data have to be used for the localization of nonlinear scatterers. Figures 10 and 12 seem to indicate a good robustness of the procedure to the presence of noise or specimen heterogeneities (also nonlinear). On the contrary, it seems that in the case of a low Q material a good reconstruction may be obtained only when in the forward propagation experiment time signals are picked in a large number of different locations, which may be a drawback for applications unless a scanning laser is available. Nevertheless, when attenuation is low, longer signals can be used, hence compensating for the number of transducers.

Using longer signals also should help increase the resolution, thanks to the more consistent mode-conversion effect between shear and longitudinal waves. That may be useful for reconstruction of the shape of the scatterer. At this stage, the focal area is strongly determined by the kind of signal used in the forward propagation problem. A more careful analysis has to be carried on about this issue.

ACKNOWLEDGMENTS

[1] M. Fink, "Time reversal of ultrasonic fields. Parts I, II and III," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–592 (1992).

[2] M. Fink, "Time reversed acoustics," Phys. Today **50**, 34–40 (1997).

[3] M. Fink, "Time-reversal mirrors," J. Phys. D **26**, 1333–1350 (1997).

[4] A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high-order multiple scattering," Phys. Rev. Lett. **75**, 4206–4209 (1995).

[5] G. A. Melkov, V. I. Vasyuchka, Yu. V. Kobljanskyj, and A. N. Slavin, "Wave-front reversal in a medium with inhomogeneities and an anisotropic wave spectrum," Phys. Rev. B **70**, 224407 (2004).

[6] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J.-L. Thomas, and F. Wu, "Time-reversed acoustics," Rep. Prog. Phys. **63**, 1933–1995 (2000).

[7] P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time-reversal acoustics," J. Acoust. Soc. Am. **111**, 230–248 (2002).

[8] G. Draeger, G. Cassereau, and D. Fink, "Theory of the time-reversal process in solids," J. Acoust. Soc. Am. **102**, 1289–1295 (1997).

[9] P. P. Delsanto, P. A. Johnson, M. Scalerandi, and J. A. TenCate, "LISA simulations of time-reversed acoustic and elastic wave experiments," J. Phys. D **35**, 3145–3152 (2002).

[10] A. M. Sutin, J. A. TenCate, and P. A. Johnson, "Single-channel time reversal in elastic solids," J. Acoust. Soc. Am. **116**, 2779–2784 (2004).

[11] M. Tanter, J.-L. Thomas, F. Coulouvrat, and M. Fink, "Time reversal invariance of nonlinear acoustic wave propagation in weakly viscous media," *IEEE Ultrasonics Symposium* (IEEE, Piscataway, 1999), pp. 419–422.

[12] K. B. Cunningham, M. F. Hamilton, A. P. Brysev, and L. M. Krutyansky, "Time reversed sound beams of finite amplitude," J. Acoust. Soc. Am. **109**, 2668–2674 (2001).

[13] D. R. Jackson and D. R. Dowling, "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**, 171–181 (1991).

[14] L. J. Huang, M. Fehler, and R. S. Wu, "Extended local Born Fourier migration method," Geophys. J. **64**, 1524–1534 (1999).

[15] J. Tromp, C. Tape, and Q. Liu, "Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels," Geophys. J. Int. **160**, 195–216 (2005).

[16] J. L. Thomas and M. A. Fink, "Ultrasonic beam focusing through tissue inhomogeneities with a time reversal mirror: Application to transskull therapy," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1122–1129 (1996).

[17] J.-L. Thomas, F. Wu, and M. Fink, "Time reversal focusing applied to lithotripsy," Ultrason. Imaging **18**, 106–121 (1996).

[18] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).

[19] P. Roux, B. Roman, and M. Fink, "Time-reversal in an ultrasonic waveguide," Appl. Phys. Lett. **70**, 1811–1813 (1998).

[20] Y. Xu and L. V. Wang, "Time reversal and its application to tomography with diffracting sources," Phys. Rev. Lett. **92**, 339021–339024 (2004).

[21] C. Prada *et al.*, "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," J. Acoust. Soc. Am. **99**, 2067–2077 (1996).

[22] C. Prada, E. Kerbrat, D. Cassereau, and M. Fink, "Time reversal techniques in ultrasonic nondestructive testing of scattering media," Inverse Probl. **18**, 1761–1773 (2002).

[23] X. D. Wang and G. L. Huang, "Identification of embedded cracks using back-propagating elastic waves," Inverse Probl. **20**, 1393–1409 (2004).

[24] V. Giurgiutiu and A. Cuc, "Embedded non-destructive evaluation for structural health monitoring, damage detection, and failure prevention," Schock Vibr. Dig. **37**, 83105 (2005).

[25] C. H. Wang, J. T. Rose, and F-K. Chang, "A computerized timereversal method for structural health monitoring," Proc. SPIE **5046**, 4858 (2003).

[26] H. W. Park, H. Sohn, K. H. Law, and C. R. Farrar, "Time reversal active sensing for health monitoring of a composite plate," J. Sound Vib. in press (2006).

[27] T. Leutenegger and J. Dual, "Detection of defects in cylindrical structures using a time reverse method and a finite-difference approach," Ultrasonics **40**, 721–725 (2002).

[28] H. W. Chun, J. T. Rose, and F-K. Chang, "A synthetic time-reversal imaging method for structural health monitoring," Smart Mater. Struct. **13**, 415–423 (2004).

[29] P. P. Delsanto and M. Scalerandi, "Modeling nonclassical nonlinearity, conditioning and slow dynamics effects in mesoscopic elastic materials," Phys. Rev. B **68**, 064107 (2003).

[30] M. Nobili and M. Scalerandi, "Temperature effects on the elastic properties of hysteretic elastic media: Modeling and simulations," Phys. Rev. B **69**, 104105 (2004).

[31] V. V. Kazakov, A. Sutin, and P. A. Johnson, "Sensitive imaging of an elastic nonlinear wave-scattering source in a solid," Appl. Phys. Lett. **81**, 646–648 (2002).

[32] R. A. Guyer and P. A. Johnson, "Nonlinear mesoscopic elasticity: Evidence for a new class of materials," Phys. Today **52**, 30–36 (1999).

[33] P. P. Delsanto, A. S. Gliozzi, M. Hirsekorn, and M. Nobili, "A 2D spring model for the simulation of ultrasonic wave propagation in nonlinear hysteretic media," Ultrasonics **44**, 279–286 (2006).

[34] R. S. Schechter, H. H. Chaskelis, R. B. Mignogna, and P. P. Delsanto, "Real-time parallel and visualization of ultrasonic pulses in solids," Science **265**, 1188–1192 (1994).

[35] P. P. Delsanto and M. Scalerandi, "A spring model for the simulation of wave propagation across imperfect interfaces," J. Acoust. Soc. Am. **104**, 2584–2591 (1998).

[36] P. P. Delsanto, R. B. Mignogna, and R. S. Schechter, "Connection machine simulation of ultrasonic wave propagation in materials II: The two dimensional case, Wave Motion **20**, 295–314 (1994).

[37] M. Bentahar, H. El Aqra, R. El Guerjouma, M. Griffa, and M. Scalerandi, "Hysteretic elasticity in damaged concrete," Phys. Rev. B **73**, 014116 (2006).

[38] M. Scalerandi, P. P. Delsanto, V. Agostini, K. Van Den Abeele, and P. A. Johnson, "Local interaction simulation approach to modeling nonclassical, nonlinear elastic behavior in solids," J. Acoust. Soc. Am. **113**, 3049–3059 (2003).

[39] R. A. Guyer, K. R. McCall, G. N. Boitnott, L. B. Hilbert Jr., and T. J. Plona, "Quantitative implementation of Preisach-Mayergoyz space to find static and dynamic elastic moduli in rock," J. Geophys. Res. **102**, 5281–5293 (1997).

[40] P. P. Delsanto, N. K. Batra, R. B. Mignogna, and M. Scalerandi, "Parallel processing simulations of the propagation of ultrasonic waves through material interfaces," *Proc. IEEE Ultras. Symposium* (IEEE, Piscataway, 1998), Vol. **2**, pp. 1129–1138.

[41] L. Borcea, G. Papanicolau, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," Inverse Probl. **18**, 1247–1279 (2002).

[42] S. K. Lehman and A. J. Devaney, "Transmission mode time-reversal super-resolution imaging," J. Acoust. Soc. Am. **113**, 2742–2753 (2003).

[43] F. K. Gruber, E. A. Marengo, and A. J. Devaney, "Time-reversal imaging with multiple signal classification considering scattering between the targets," J. Acoust. Soc. Am. **115**, 3042–3047 (2004).

[44] P. Blomgren, G. Papanicolau, and H. Zhao, "Long distance time reversal and imaging in random media," J. Acoust. Soc. Am., **110**, 2633 (2001).

[45] A. S. Gliozzi, M. Nobili, and M. Scalerandi, "Modeling and characterization of localized micro-damage through resonance measurements," J. Phys. D **39**, 3895–3903 (2006).

[46] K. Van den Abeele, S. Sutin, and P. A. Johnson, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, Part I: Nonlinear wave modulation spectroscopy (NWMS)," Res. Nondestruct. Eval. **12**, 17–30 (2000).

The partial-wave expansion for scattering in waveguides

Dalcio K. Dacol
Acoustics Division, Naval Research Laboratory, Washington, DC 20375-5350

Dilip G. Roy
*Medical Imaging Research Laboratory, Department of Radiology, The University of Utah,
Salt Lake City, Utah 84109*

The scattering of acoustic waves by objects located inside a waveguide is discussed, assuming that the scattering amplitude for the object in an extended uniform medium is known. The scattering process is described by using an expansion of the scattering amplitude in terms of spherical harmonics. An appropriate multipole decomposition of the waveguide Green's function is developed and the effective scattering amplitude in the waveguide is obtained. An important property of the effective scattering amplitude, the generalized optical theorem, is obtained and its implications for scattering in a waveguide are discussed. The scattering problem is formulated entirely and explicitly in terms of the waveguide Green's functions, which makes this approach very flexible in regard to the choice of the incident field. It also establishes the connection between propagation and scattering and allows for the independent computation of the propagation and scattering aspects of the problem. This is the main advantage of using the scattering amplitude in an extended uniform medium as an input. The connection of this work with previous work in scattering in waveguides is discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354077]

Pages: 2518–2525

I. INTRODUCTION

The scattering of acoustic waves by an object imbedded in a heterogeneous acoustic medium is the problem addressed in this paper. The heterogeneity may originate from nonuniform density and sound speed, the presence of boundaries and interfaces, or a combination of all these factors. The difficulty in solving this problem arises from the fact that one has to enforce boundary conditions both on the surface of the scattering object as well as on the interfaces and boundaries. For nonuniform media, additional difficulties arise, since only numerical solutions are usually available. However, a promising way of tackling those problems is to assume that the single object scattering problem in an extended uniform medium can be solved independently and then formulate the waveguide scattering problem in such a way as to incorporate this result into the full solution of the problem. The pseudopotential introduced by Huang and Yang[1] offers a convenient starting point since it separates the implementation of both sets of boundary conditions. With this approach, the scattering properties of the object are subsumed in a series of field-dependent source terms added to the wave equation, the solution of this equation must then satisfy proper boundary conditions on the medium's boundaries.[2]

The basic formulation of the approach adopted in this work was presented in Ref. 2. In Ref. 3 the method was further developed and used to obtain the generalized optical theorem for a scattering object embedded in an inhomogeneous medium. Its use in scattering problems in ocean acoustics is elaborated on here. The formulation is entirely based on Green's function that makes the approach independent of particular assumptions about the incident field, as all incident fields can be obtained with the Green's functions and appropriate source functions. Also the formulation is independent of any particular explicit expression for the Green's functions. An explicit operational procedure is obtained to produce the proper multipoles of the waveguide Green's functions that enter in the scattering equations. If one examines previous formulations of this problem, one notices that a great deal of effort is involved in computing transformations from cylindrical to spherical coordinates and basis functions.[4,10] While those calculations are unavoidable when one has to carry out computations they are not necessary when obtaining the basic equations as demonstrated in this article and in Refs. 2 and 3. This simplification allowed the authors to discuss the energy flux in the scattering process in a relative simple and transparent way culminating in the derivation of the generalized optical theorem taking into account all the multiple interactions of the scattered acoustic field with boundaries and other spatially varying acoustic properties of the waveguide. The only other discussion of the optical theorem in underwater acoustics known to the authors, namely Ref. 13, neglects those interactions.

There are other approaches to this problem that are explicit designed with numerical computations as the goal and thus differ in spirit from the work presented here. A particularly interesting work is the one in Ref. 5 where multiple boundary interactions for a cylindrical object in a waveguide is discussed through an ingenious numerical approach. Similarly, the problem of acoustic scattering by an elastic spherical shell near a perfectly reflecting plane boundary was discussed in Ref. 6 by the method of images. In Ref. 7 one finds another interesting approach to this problem, which is also pragmatically oriented toward numerical computations. In this paper we emphasize a general formulation of the prob-

lem, delaying as much as possible the use of explicit expressions for the Green's functions, the scattering amplitudes, etc. This allows further analysis, such as obtaining the optical theorem. This, of course, does not preclude numerical calculations but permits one to explore how far one can go before resorting to computations.

The general formulation is presented in Sec. II. In Sec. III we discuss the generalized optical theorem. Since all necessary derivations are similar to those in Refs. 2 and 3, they are not repeated here. In Sec. IV we present results pertaining to scattering in an ideal waveguide (computations for more realistic waveguides will be presented in a future work). In order to illustrate the effect of boundary interactions, these results are compared with results for the same problem obtained using Ingenito's[4] approximate formulation of the scattering problem. Finally, in Sec. V we summarize and further discuss the results.

II. SCATTERING IN AN INHOMOGENEOUS MEDIUM

In this section the general results obtained in Ref. 2 will be presented in the context of an ocean acoustic waveguide. As the derivation parallels that of Ref. 2, it will be omitted here. In this formulation the scattering in an extended and uniform medium with sound speed c of a cw plane wave with frequency $\omega/2\pi$ by an arbitrary scattering object is characterized by an operator T given by the following projection of the scattering amplitude into spherical harmonics:

$$T_{lm,l'm'}(k) = \frac{i^{l-l'}}{4\pi} \int d\hat{\mathbf{p}} \, Y_{lm}^*(\hat{\mathbf{p}}) \int d\hat{\mathbf{q}} \, Y_{l'm'}(\hat{\mathbf{q}}) f(\hat{\mathbf{p}}, \hat{\mathbf{q}}). \quad (1)$$

In Eq. (1) $f(\hat{\mathbf{p}}, \hat{\mathbf{q}})$ is the scattering amplitude defined in the usual way in terms of the asymptotic limit for $r \rightarrow \infty$ of the scattered wave field created by the scattering of an incident plane wave by the object.[8] Note that the expression in Eq. (1) differs from the one in Ref. 2 by a wavenumber factor $k = \omega/c$.

The procedure developed in Ref. 2 can be used to obtain equations describing scattering in a general case and it can be formulated without a specific expression for the Green's function in the waveguide. This Green's function obeys Bergmann's extension of the Helmholtz equation[9] to a nonuniform fluid plus appropriate boundary conditions:

$$\rho(\mathbf{r}) \, \mathbf{\nabla} \cdot \left(\frac{1}{\rho(\mathbf{r})} \, \mathbf{\nabla} G_0(\mathbf{r}, \mathbf{r}_0) \right) + k(\mathbf{r})^2 G_0(\mathbf{r}, \mathbf{r}_0) = \delta(\mathbf{r} - \mathbf{r}_0).$$

$$(2)$$

In order to use the operator T as an input in the computation of scattering in an arbitrary medium, one must assume that in a neighborhood fully containing the object in the arbitrary medium the acoustic wavenumber and the density can be assumed to be locally uniform. Thus, for $\mathbf{r}$ and $\mathbf{r}_0$ in this region, that is for $d+\varepsilon > |\mathbf{r} - \mathbf{r}_T| > d$ and $d+\varepsilon > |\mathbf{r}_0 - \mathbf{r}_T| > d$, where d is the largest linear dimension of the object and ε is positive (perhaps very small), $G_0(\mathbf{r}, \mathbf{r}_0)$ can be written as

$$G_0(\mathbf{r}, \mathbf{r}_0) = g_T(\mathbf{r} - \mathbf{r}_0) + G_0^{\text{NS}}(\mathbf{r}, \mathbf{r}_0),$$

$$(3)$$

$$g_T(\mathbf{r}) = -\frac{e^{ik_T r}}{4\pi r}, \quad k_T = k(\mathbf{r}_T).$$

That is, the Green's function inside a region of uniform density and sound speed can be split into two parts: a singular one, which is just the Green's function in the equivalent unbounded medium, and a nonsingular part enforcing the boundary conditions.

The techniques employed in Ref. 2 lead to an equation for the waveguide Green's function when a scattering object is present in terms of a set of partial wave coefficients that are determined by a linear algebraic equation for the new coefficients. Thus the Green's function for a waveguide with a scattering object inside can be written as

$$G(\mathbf{r}, \mathbf{r}_0) = G_0(\mathbf{r}, \mathbf{r}_0) - \sum_{lm} \sum_{l'm'} G_{lm}^{\text{out}}(\mathbf{r}, \mathbf{r}_T) T_{lm,l'm'} a_{l'm'}. \quad (4)$$

For nonspherically symmetric scatterers the above equation is valid for $\mathbf{r}$ outside the enclosing sphere, that is, it is valid when the point $\mathbf{r}$ is outside of a sphere whose diameter is equal to the largest linear dimension of the scattering object.

The coefficients a_{lm} obey the following equation:

$$a_{lm} = G_{lm}^{\text{inc}} - \sum_{l'm'} \sum_{l''m''} K_{lm,l'm'}(\mathbf{r}_T) T_{l'm',l''m''} a_{l''m''}. \quad (5)$$

The new quantities introduced in the above equations are as follows:[2,3]

$$G_{lm}^{\text{out}}(\mathbf{r}, \mathbf{r}_T) = D_{lm}^*(\mathbf{\nabla}') G_0(\mathbf{r}, \mathbf{r}')|_{\mathbf{r}'=\mathbf{r}_T} \quad \text{for } \mathbf{r} \neq \mathbf{r}_T, \quad (6)$$

$$G_{lm}^{\text{inc}}(\mathbf{r}_T, \mathbf{r}_0) = D_{lm}(\mathbf{\nabla}) G_0(\mathbf{r}, \mathbf{r}_0)|_{\mathbf{r}=\mathbf{r}_T} \quad \text{for } \mathbf{r}_0 \neq \mathbf{r}_T, \quad (7)$$

and

$$K_{lm,l'm'}(\mathbf{r}_T) = D_{lm}(\mathbf{\nabla}) D_{l'm'}^*(\mathbf{\nabla}') G_0^{\text{NS}}(\mathbf{r}, \mathbf{r}')|_{\mathbf{r}=\mathbf{r}'=\mathbf{r}_T}. \quad (8)$$

In the above equations $D_{lm}(\mathbf{\nabla})$ is a differential operator, a polynomial of degree l in the partial derivative operators ∂_x, ∂_y, and ∂_z, defined as follows:[3]

$$D_{lm}(\mathbf{\nabla}) = \frac{(2l+1)!!}{k_T^l} \frac{1}{l!} \int d\hat{\mathbf{s}} \, Y_{lm}(\hat{\mathbf{s}})^* (\hat{\mathbf{s}} \cdot \mathbf{\nabla})^l$$

$$= \frac{4\pi i^l}{k_T^l} \frac{(-1)^m}{2^l l!} N_{lm} (p_x - ip_y)^m \partial_u^{l+m} (u^2 - p^2)^l \Big|_{u=p_z},$$

$$(9)$$

where $\mathbf{p} = -i\mathbf{\nabla}$, $p^2 = -\nabla^2$, and N_{lm} is the normalization factor for the spherical harmonics. The resulting equations, Eqs. (4) and (5), show that, in the presence of boundaries and/or a heterogeneous medium, there is a coupling among the scattered partial waves, even for spherically symmetrical scatterers. One should also notice that the only quantity related to the scattering target that appears in the coupling matrix, Eq. (8), is its position relative to the boundaries. This indicates that no matter what the nature of the target is, this coupling matrix determines the coupling among the scattered partial waves induced by the inhomogeneities of the medium, such as spatially varying acoustic properties, interfaces, and boundaries.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

D. K. Dacol and D. G. Roy: Scattering in waveguides   2519

The equation for the Green's function, including scattering by an object at $\mathbf{r} = \mathbf{r}_T$, can now be written as

$$G(\mathbf{r}, \mathbf{r}_0) = G_0(\mathbf{r}, \mathbf{r}_0) - \sum_{lm} \sum_{l'm'} D_{lm}^*(\boldsymbol{\nabla}')$$
$$\times |G_0(\mathbf{r}, \mathbf{r}')| \mathbf{T}_{lm,l'm'}(\mathbf{r}_T) D_{l'm'}(\boldsymbol{\nabla}'')$$
$$\times |G_0(\mathbf{r}'', \mathbf{r}_0)|_{\mathbf{r}'=\mathbf{r}''=\mathbf{r}_T}, \tag{10}$$

where the effective scattering operator, $\mathbf{T}(\mathbf{r}_T)$, is given by

$$\mathbf{T}(\mathbf{r}_T) = T(k_T) - \mathbf{T}(\mathbf{r}_T)K(\mathbf{r}_T)T(k_T), \quad \text{or} \tag{11}$$

$$\mathbf{T}(\mathbf{r}_T) = T(k_T)[I + K(\mathbf{r}_T)T(k_T)]^{-1}.$$

The scattering term in Eq. (10) can be interpreted as stating that a multipole decomposition of the incident field, $D_{l'm'}(\boldsymbol{\nabla}'')G_0(\mathbf{r}'', \mathbf{r}_0)|_{\mathbf{r}''=\mathbf{r}_T}$, is scattered into a multipole decomposition of the outgoing scattered field, $D_{lm}^*(\boldsymbol{\nabla}')G_0(\mathbf{r}, \mathbf{r}')|_{\mathbf{r}'=\mathbf{r}_T}$. In the case of a layered medium, this result, apart from differences in notation, coincides, formally, with that obtained by Sammelmann and Hackman.[10] However, the formulation presented here in terms of the waveguide Green's function with an explicitly operational procedure to obtain the proper multipole expansion seems to be more general, although equivalent to other formulations that start with an explicit expansion of the acoustic field in cylindrical coordinates and then carry out a projection into spherical harmonics.[4,10].

As an example to illustrate the structure of the above equations, let us consider the case where the wavelength of the incident field on the object is much larger than a typical dimension of the object. In this case it is a reasonable approximation to consider the object an isotropic scatterer where only the terms with $l=0$ contribute to the scattered field. One obtains in this case the following expression for the Green's function, including the scattered field:

$$G(\mathbf{r}, \mathbf{r}_0) = G_0(\mathbf{r}, \mathbf{r}_0) - \frac{4\pi f}{1 + 4\pi f G_0^{NS}(\mathbf{r}_T, \mathbf{r}_T)}$$
$$\times G_0(\mathbf{r}, \mathbf{r}_T)G_0(\mathbf{r}_T, \mathbf{r}_0), \tag{12}$$

where $f = T(k_T)_{00,00}$ is the isotropic scattering amplitude. Thus, one sees that in the inhomogeneous medium the effective scattering amplitude is

$$f_{\text{eff}} = \frac{f}{1 + 4\pi f G_0^{NS}(\mathbf{r}_T, \mathbf{r}_T)}. \tag{13}$$

The denominator in Eqs. (12) and (13) is a consequence of the multiple interactions between the scattered field and the boundaries and other spatially varying acoustic properties of the waveguide. The most obvious consequence of Eq. (13) is that the effective scattering amplitude varies with the position of the scattering object. The closer the object is to a boundary or interface the stronger this effect will be. Another consequence is that even if the scattering amplitude in the extended and uniform medium is not resonant, it is possible for the effective scattering amplitude to be resonant. If f is resonant, then the resonant frequency will be shifted in f_{eff}; see, for example, Ref. 11. The implications of this reso-

nance shift for bubbles near a pressure release surface are discussed in Ref. 12.

III. THE GENERALIZED OPTICAL THEOREM

An important relation satisfied by the effective scattering operator defined in Eq. (11) can be obtained by examining the energy flux across a surface that encloses the scattering object. A lucid discussion of energy fluxes in a scattering context is presented in Ref. 13. Those authors use Ingenito's[4] approximate description of scattering in a waveguide that neglects multiple interactions of the scattered field with the waveguide's boundaries, interfaces, and other spatially varying acoustic properties. Thus, their discussion is incomplete in this regard. The discussion presented in this section uses the full scattering description developed in Sec. II and addresses the impact of those multiple interactions on the energy conservation in the scattering process. A detailed derivation of the generalized optical theorem for scattering in an inhomogeneous medium is presented in Ref. 3 and, with very minor changes, that derivation is applicable to scattering in an oceanic waveguide. Unless explicitly stated, the discussion follwing is not restricted to a layered medium.

Consider the case of a nonenergy-absorbing medium and an object that also does not absorb energy. Then the total energy flux across a closed surface that surrounds the scattering object but excludes the point source is zero since all the incident acoustic energy eventually leaves the volume enclosed by the surface. The incident field carries part of this energy and part of it is in the scattered field but none is dissipated into heat or other forms of energy. The total energy flux across the closed surface is

$$F_T = \frac{1}{2i\omega} \int d\mathbf{S} \cdot \left(G(\mathbf{r}, \mathbf{r}_0)^* \frac{1}{\rho(\mathbf{r})} \boldsymbol{\nabla} G(\mathbf{r}, \mathbf{r}_0) \right.$$
$$\left. - G(\mathbf{r}, \mathbf{r}_0) \frac{1}{\rho(\mathbf{r})} \boldsymbol{\nabla} G(\mathbf{r}, \mathbf{r}_0)^* \right). \tag{14}$$

Using the divergence theorem and the equation satisfied by the full Green's function $G(\mathbf{r}, \mathbf{r}_0)$, Eq. (10), one obtains

$$F_T = \frac{1}{2i\omega\rho(\mathbf{r}_T)} G_{\text{inc}}^{\dagger}[\mathbf{T}^{\dagger} - \mathbf{T} + 2ik_T\mathbf{T}^{\dagger}\mathbf{T} + \mathbf{T}^{\dagger}$$
$$\times (K^{\dagger} - K)\mathbf{T}]G_{\text{inc}}. \tag{15}$$

In the above equation, the $G_{lm}^{\text{inc}}(\mathbf{r}_T, \mathbf{r}_0)$, defined in Eq. (7), are considered to be components (labeled by the partial wave indexes lm) of a vector G_{inc} and, $\mathbf{T}^{\dagger}$ is the Hermitian conjugate of $\mathbf{T}$, that is $(\mathbf{T}^{\dagger})_{lm,l'm'} = (\mathbf{T}_{l'm',lm})^*$. Since G_{inc} is arbitrary (the source position is arbitrary), in order for the total flux to be equal to zero the expression between square brackets on the right-hand side of Eq. (15) must also be equal to zero. Thus, the generalized optical theorem for the effective scattering operator can be stated as the following relationship between the components of $\mathbf{T}$:

$$2ik_T\mathbf{T}^{\dagger}\mathbf{T} = \mathbf{T} - \mathbf{T}^{\dagger} + \mathbf{T}^{\dagger}(K - K^{\dagger})\mathbf{T}. \tag{16}$$

In a similar way it can be shown that for a nonenergy-absorbing scattering object the scattering operator T obeys a similar generalized optical theorem:[14]

$$2ik_T T^\dagger T = T - T^\dagger. \tag{17}$$

Comparing Eqs. (16) and (17), one sees that the formal difference between them is related to the operator K. This operator accounts for multiple interactions between the scattering object and the boundaries and other spatial variations of the waveguide acoustic properties. The multiple bounces between the object and a boundary are simple to visualize, but a position dependence of the sound speed or density will have similar effects. For example, suppose the object is in a sound channel corresponding to a minimum of the sound speed as a function of depth. Some of the energy scattered off the object will be redirected toward the object by refraction and rescattered by the object, cumulative repetition of this process also leads to the presence of the operator K in the expression for the effective scattering operator. The explicit presence of K in Eq. (16) indicates that those multiple scattering processes trap a fraction of the energy incident on the object. If the object is assumed to be in a homogeneous layer, then the quantitative importance of K depends on the ratio between a typical dimension of the object and its distance from the layer boundaries, the greater this ratio the larger the impact of the multiple scattering processes that K incorporates. This is related to the phenomena of trapped modes and complex resonances in waveguides containing a scattering obstacle (see Ref. 15 for a detailed discussion of trapped modes and further references on this matter).

As in Ref. 13, let us introduce the quantities $\Phi_{inc}(\mathbf{r})$, the incident field, and $\Phi_{sct}(\mathbf{r})$, the scattered field. The velocity vectors associated with those fields are defined as $\mathbf{V}_\alpha = (1/i\omega\rho(\mathbf{r}))\nabla\Phi_\alpha(\mathbf{r})$. The total flux can be written as $F_T = F_{inc} - E + W_{sct}$. The incident flux, F_{inc}, is zero, since the source is outside the volume delimited by the closed surface, W_{sct} is the scattered energy flux involving only $\Phi_{sct}(\mathbf{r})$, and E is the extinction term that involves the negative of the contribution to the total flux due to the interference of the incident and scattered fields:

$$W_{sct} = \mathrm{Re}\left\{ \oint d\mathbf{S}\cdot\mathbf{V}_{sct}^*\Phi_{sct} \right\},$$

$$E = -\mathrm{Re}\left\{ \oint d\mathbf{S}\cdot[\mathbf{V}_{sct}^*\Phi_{inc} + \mathbf{V}_{inc}^*\Phi_{sct}] \right\}. \tag{18}$$

Procedures similar to the ones that lead to Eq. (17) yield

$$W_{sct} = \frac{1}{2i\omega\rho(\mathbf{r}_T)}G_{inc}^\dagger[2ik_T\mathbf{T}^\dagger\mathbf{T} + \mathbf{T}^\dagger(K^\dagger - K)\mathbf{T}]G_{inc}, \tag{19}$$

$$E = \frac{1}{2i\omega\rho(\mathbf{r}_T)}G_{inc}^\dagger[\mathbf{T} - \mathbf{T}^\dagger]G_{inc}.$$

Thus, the generalized optical theorem states that, in the absence of energy absorption by the object and the medium, the scattered power is equal to the power subtracted from the incident field by its interference with the scattered field. In the case of an energy-absorbing object, equations such as (15) and (19) still hold but (16) does not as Eq. (17) is no longer true since a term due to energy absorption by the scattering object is missing. In this case the expression for

the power extinguished from the incident field is still given by the corresponding equation in Eq. (19). In the case of an energy-absorbing medium the procedures that yielded Eqs. (15) and (19) generate extra terms that have to do with the power absorbed by the medium in the volume enclosed by the closed surface across which the fluxes were computed. It is clear that it is not empirically possible to separate extinction effects due to scattering and absorption by the scattering object from the effects of energy absorption in the medium.

IV. NUMERICAL EXAMPLES

In this section, results for scattering in a simple waveguide using the approach developed in Sec. II are compared with results obtained using Ingenito's approximation,[4] which neglects the multiple interactions between the scatterer and the boundaries. An ideal waveguide is considered here to be a planar waveguide with a uniform fluid medium filling the space between the plane boundaries, the upper boundary being a pressure-release surface while the bottom boundary is acoustically hard. While the acoustically hard boundary condition is not very realistic for ocean waveguides, it is used here to provide an illustration of the relationship between boundary conditions and the multiple scattering effect. A detailed study of the impact of more realistic bottoms on the multiple scattering effect will be delegated to a future article.

The method of images allows the expression of the Green's function for this ideal waveguide as a superposition of the free space Green's functions for the acoustic medium. If the partial wave expansion for the scattering matrix of the object is also known, this simplifies considerably the computations. In the following the scattering object is taken to be a sphere with a pressure-release surface that has a simple and well-known expression for its scattering amplitude in terms of partial waves.

In the case of the waveguide, the Green's function can also be expressed as a superposition of Green's functions for the unbounded medium:[16]

$$G_0(\mathbf{r},\mathbf{r}_0) = \sum_{n=-\infty}^{\infty} (-1)^n[g(\mathbf{r}-\mathbf{r}_0^+ - 2nh\hat{\mathbf{e}}_z) - g(\mathbf{r}-\mathbf{r}_0^- - 2nh\hat{\mathbf{e}}_z)], \tag{20}$$

where $\mathbf{r}_0^\pm = \mathbf{R}_0 \pm z_0\hat{\mathbf{e}}_z$. In particular, one has for the nonsingular part of the Green's function [defined in Eq. (3)],

$$G_0^{NS}(\mathbf{r},\mathbf{r}_0) = -g(\mathbf{r}-\mathbf{r}_0^-) + \sum_{n\neq 0} (-1)^n[g(\mathbf{r}-\mathbf{r}_0^+ - 2nh\hat{\mathbf{e}}_z) - g(\mathbf{r}-\mathbf{r}_0^- - 2nh\hat{\mathbf{e}}_z)]. \tag{21}$$

The Green's function in an unbounded medium with wavenumber k has a well-known form:

$$g(\mathbf{r}-\mathbf{r}') = -\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} = \frac{1}{4\pi^2}\int d\mathbf{K}\frac{e^{i\mathbf{K}\cdot(\mathbf{R}-\mathbf{R}')+ik_z(K)|z-z'|}}{2ik_z(K)}, \tag{22}$$

where the capital letters indicate vectors in the xy plane and

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

D. K. Dacol and D. G. Roy: Scattering in waveguides   2521

$$k_z(K) = \sqrt{k^2 - K^2} \text{ if } k \geqslant K \text{ and } k_z(K)$$
$$= i\sqrt{K^2 - k^2} \text{ if } K > k. \tag{23}$$

A few useful quantities are obtained from the partial wave expansion of the free space Green's function, (in the following $k_T = k$, the wavenumber in the homogeneous fluid-filled half-space):

$$D_{lm}(\nabla)g(\mathbf{r} - \mathbf{r}_0)|_{\mathbf{r}=\mathbf{r}_T} = -ikh_l^{(1)}(ks)Y_{lm}^*(\hat{\mathbf{s}})|_{\mathbf{s}=\mathbf{r}_0-\mathbf{r}_T},$$
$$\text{for } \mathbf{r}_0 \neq \mathbf{r}_T. \tag{24}$$

Similarly, one obtains

$$D_{lm}^*(\nabla')g(\mathbf{r} - \mathbf{r}')|_{\mathbf{r}'=\mathbf{r}_T} = -ikh_l^{(1)}(ks)Y_{lm}(\hat{\mathbf{s}})|_{\mathbf{s}=\mathbf{r}-\mathbf{r}_T},$$
$$\text{for } \mathbf{r} \neq \mathbf{r}_T. \tag{25}$$

Another useful quantity is the reduced coupling operator $K^g(\mathbf{r}_T)$, with matrix elements given by the following expression:

$$K_{lm,l'm'}^g(\mathbf{r}_T) = -D_{lm}(\nabla)D_{l'm'}^*(\nabla')g(\mathbf{r} - \mathbf{r}')|_{\mathbf{r}=\mathbf{r}_T^+, \mathbf{r}'=\mathbf{r}_T^-}. \tag{26}$$

In order to obtain an expression for $K_{lm,l'm'}^g(\mathbf{r}_T)$, the integral representation for the Green's function in Eq. (22) will be used. First, one should notice that

$$D_{lm}(\nabla)\exp(i\mathbf{q} \cdot \mathbf{r}) = 4\pi\left(\frac{iq}{k}\right)^l Y_{lm}(\hat{\mathbf{q}})^*\exp(i\mathbf{q} \cdot \mathbf{r}). \tag{27}$$

Notice that the above result is entirely algebraic. That is, it is valid even if $\mathbf{r}$ or $\mathbf{q}$ or both have complex components. In the case of complex components, one should write $(-1)^m Y_{l-m}(\hat{\mathbf{q}})$ instead of $Y_{lm}(\hat{\mathbf{q}})^* = (-1)^m Y_{l-m}(\hat{\mathbf{q}}^*)$, that is, the complex conjugation in Eq. (27) does not apply to the argument of the spherical harmonic. However, for conciseness of notation, the form in Eq. (27) is preferable. In the same way one also obtains

$$D_{lm}^*(\nabla)\exp(-i\mathbf{q} \cdot \mathbf{r}) = 4\pi\left(-\frac{iq}{k}\right)^l Y_{lm}(\hat{\mathbf{q}})\exp(-i\mathbf{q} \cdot \mathbf{r}). \tag{28}$$

Thus, for the coupling operator one has this expression

$$K_{lm,l'm'}^g(\mathbf{r}_T) = -(4\pi)^2(i)^l(-i)^{l'}$$
$$\times \int \frac{d\mathbf{K}}{(2\pi)^2}Y_{lm}(\hat{\mathbf{k}}^+)^*Y_{l'm'}(\hat{\mathbf{k}}^-)\frac{e^{2ik_z(K)z_T}}{2ik_z(K)}, \tag{29}$$

where $\mathbf{k}^{\pm} = \mathbf{K} \pm k_z(K)\hat{\mathbf{e}}_z$. The above integral can be carried out, and one obtains

$$K_{lm,l'm'}(\mathbf{r}_T) = \delta_{m,m'}L_{l,l'}^m(2z_T), \tag{30}$$

where

$$L_{l,l'}^m(z) = 4\pi(i)^l(-i)^{l'}N_{lm}N_{l'm}\frac{e^{ikz}}{z}$$
$$\times \sum_{p=0}^{l+l'}\left(-\frac{1}{ikz}\right)^p \partial_x^p[P_l^m(x)P_{l'}^m(-x)]|_{x=1}. \tag{31}$$

The quantity N_{lm} is the normalization factor of the spherical harmonics, that is, $Y_{lm}(\hat{\mathbf{r}}) = N_{lm}P_l^m(\cos\hat{\theta})e^{im\varphi}$. The finite sum in Eq. (31) is due to the fact that the product of Legendre functions on the right-hand side of this equation is a polynomial of degree $l+l'$. That K is diagonal in m and m' is a consequence of the cylindrical symmetry of the environment. One notices that $L_{l,l'}^m(z)$ depends only on $|m|$. This dependence on $|m|$ is due to the fact that in the presence of an acoustic point source and a scattering sphere the original cylindrical symmetry is replaced by reflection symmetry across the plane that contains the source and sphere positions and is perpendicular to the bounding surface plane. It follows from Eq. (33) that $l \geqslant |m|$ and $l' \geqslant |m'|$ thus $L_{l,l'}^m(z) \equiv 0$ unless $l_S \geqslant |m|$, where l_s is the smallest of l and l'.

Thus the results in Eqs. (24), (25), and (30) considerably simplify solving Eqs. (4) and (5) for the case of acoustic scattering by a sphere in the ideal waveguide. Thus one obtains the following expressions:

$$G_{lm}^{inc}(\mathbf{r}_T,\mathbf{r}_0) = -ik\sum_{n=-\infty}^{\infty}(-1)^n[h_l^{(1)}(ks)$$
$$\times Y_{lm}^*(-\hat{\mathbf{s}})|_{\mathbf{s}=\mathbf{r}_T-\mathbf{r}_0^+-2nh\hat{\mathbf{e}}_z} - h_l^{(1)}(ks)$$
$$\times Y_{lm}^*(-\hat{\mathbf{s}})|_{\mathbf{s}=\mathbf{r}_T-\mathbf{r}_0^--2nh\hat{\mathbf{e}}_z}], \tag{32}$$

$$G_{lm}^{out}(\mathbf{r},\mathbf{r}_T) = -ik\sum_{n=-\infty}^{\infty}(-1)^n[h_l^{(1)}(ks)Y_{lm}(\hat{\mathbf{s}})|_{\mathbf{s}=\mathbf{r}-\mathbf{r}_T^+-2nh\hat{\mathbf{e}}_z}$$
$$- (-1)^{l+m}h_l^{(1)}(ks)Y_{lm}(\hat{\mathbf{s}})|_{\mathbf{s}=\mathbf{r}-\mathbf{r}_T^--2nh\hat{\mathbf{e}}_z}], \tag{33}$$

$$K_{lm,l'm'}(\mathbf{r}_T) = \delta_{m,m'}\left\{L_{l,l'}^m(2z_T) - \sum_{n\neq 0}(-1)^n \times [L_{l,l'}^m(-2nh) - L_{l,l'}^m(2z_T - 2nh)]\right\}. \tag{34}$$

The above equations plus the expression for the free space scattering operator for the object allows one to solve Eq. (5) for a_{lm} and then use Eq. (4) to obtain the Green's function for the waveguide in the presence of a scattering object. In the example discussed here, the scattering object is a pressure-release sphere of radius a and the scattering operator for the sphere can be written as

$$T_{lm,l'm'} = f_l\delta_{l,l'}\delta_{m,m'}, \quad f_l = \frac{i}{k}\frac{j_l(ka)}{h_l^{(1)}(ka)}. \tag{35}$$

The fact that in this example the scattering object is a sphere simplifies the structure of the algebraic equations for a_{lm}, since the equations for different values of m are not coupled as both $K_{lm,l'm'}(\mathbf{r}_T)$ and $T_{lm,l'm'}$ are diagonal in m. The linear algebraic equations for a_{lm} are solved numerically by trun-
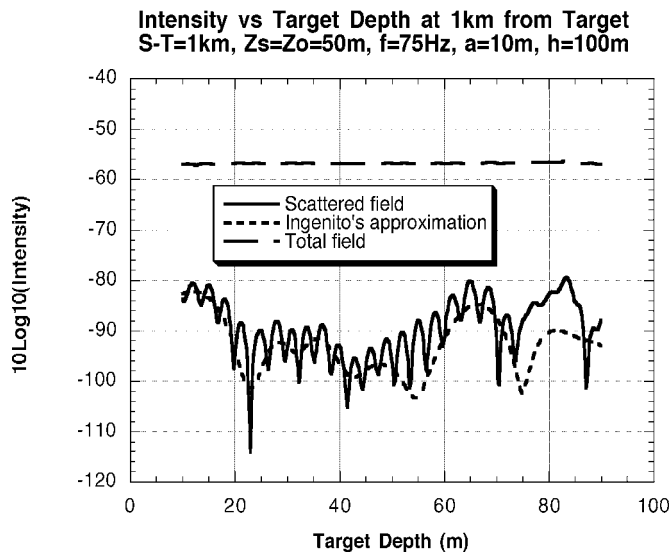
FIG. 1. Scattering in an ideal waveguide by a pressure-release, 10 m radius sphere. The horizontal distance from the scattering sphere to either the acoustic point source or the observation point is 1 km. The observation point and the source are both at a depth of 50 m in a 100 m deep waveguide. The depth of the center of the sphere varies between 10 and 90 m. The cw point source has a frequency of 75 Hz. This figure illustrates the variation of the scattered and total fields as the depth of the scattering sphere changes.

cating the partial wave expansion. The maximum value for l is taken as $l_{max} = \text{Int}(ka) + 5$, which reflects the fact that in the partial wave expansion for scattering by a sphere the contributions from terms with $l > ka$ are negligible. The total number of partial wave coefficients a_{lm} is $l_{max}(l_{max} + 2)$.

Results are presented in Figs. 1–6 which show the outcome of calculations for acoustic scattering in an ideal oceanic waveguide by a pressure-release, 10 m radius, sphere. In Figs. 1–6 the horizontal distance from the scattering sphere to either the acoustic point source or the observation point is 1 km. The source and the target are both at a depth of 50 m in a 100 m deep waveguide. The depth of the center
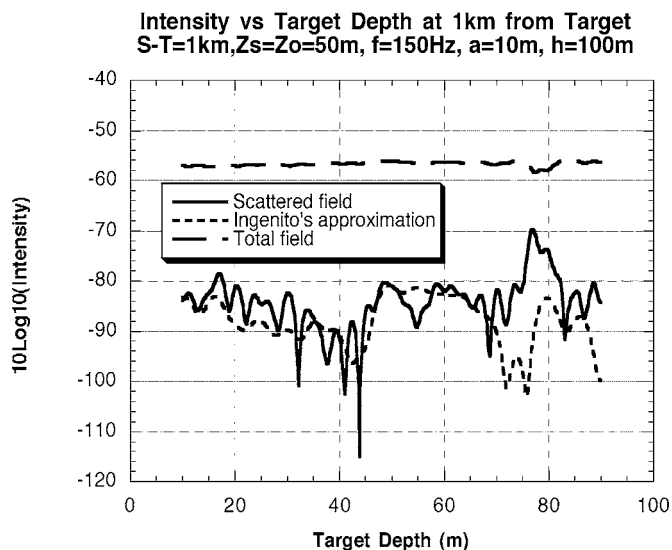
FIG. 2. This figure is similar to Fig. 1, except that in this figure the cw point source has a frequency of 150 Hz. This figure illustrates the variation of the scattered and total fields as the depth of the scattering sphere changes.
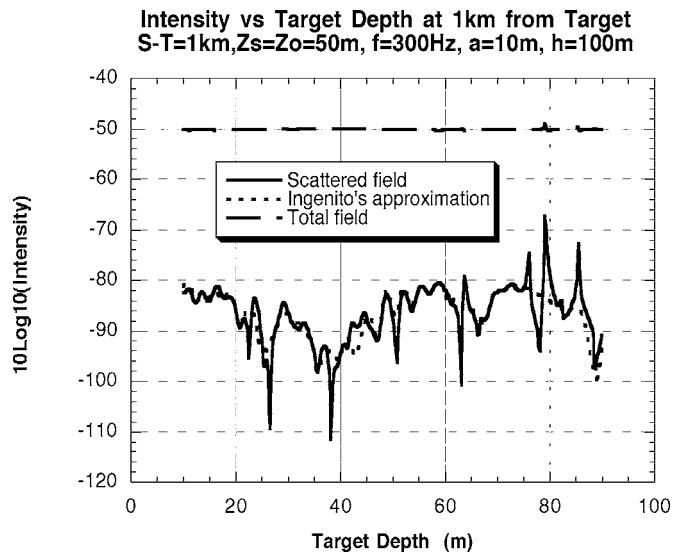
FIG. 3. This figure is similar to Figs. 1 and 2, except that in this figure the cw point source has a frequency of 300 Hz. This figure illustrates the variation of the scattered and total fields as the depth of the scattering sphere changes.

of the sphere varies between 10 and 90 m. These figures illustrate the variation of the scattered and total fields as the depth of the scattering sphere changes. The scattered field computed taking into account multiple interactions with the boundaries is compared with a computation that neglects those interactions, an approximation introduced by Ingenito.[4] There were substantial differences between the two results in those examples. In particular, notice the different impacts of the surface and bottom boundary conditions on the multiple scattering effect. Clearly the pressure-release condition at the surface has a weaker impact than acoustically hard condition
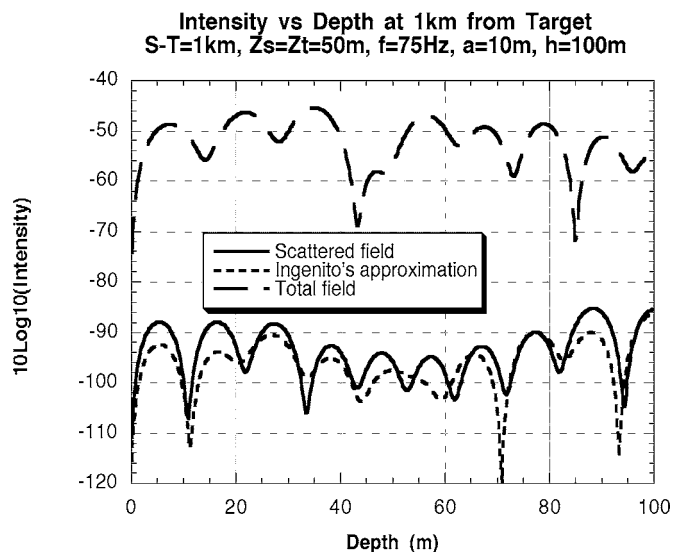
FIG. 4. Scattering in an ideal waveguide by a pressure-release, 10 m radius sphere. The horizontal distance from the scattering sphere to either the acoustic point source or the observation point is 1 km. The source and the scattering sphere are both at a depth of 50 m in a 100 m deep waveguide. The depth of the observation point varies between 0 (surface) and 100 m (bottom). The cw point source has a frequency of 75 Hz. This figure illustrates the variation of the scattered and total fields as the depth of the observation point changes.
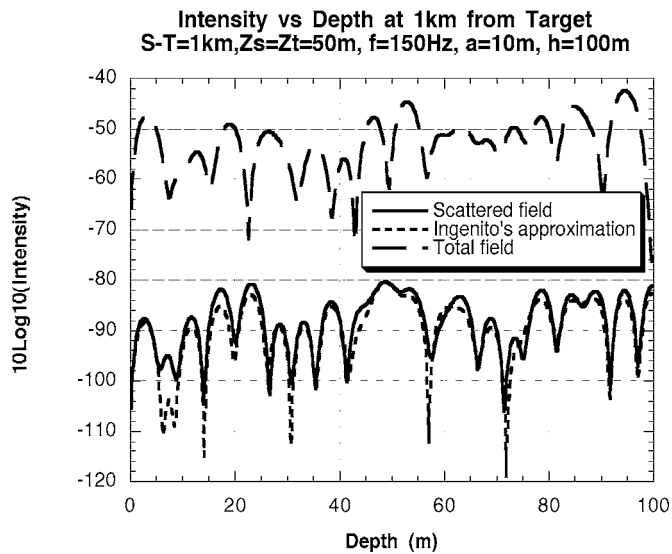
FIG. 5. This figure is similar to Fig. 4, except that in this figure the cw point source has a frequency of 150 Hz. This figure illustrates the variation of the scattered and total fields as the depth of the observation point changes.

at the bottom. Those differences were larger at lower frequencies than at higher ones, which is a consequence of the fact that the matrix elements of the coupling operator K decrease with increasing frequency mostly due to the rapid variations of the oscillating exponential factor in Eq. (31). In Figs. 3–6, the source and the scattering sphere are both at a depth of 50 m in a 100 m deep waveguide. The depth of the observation point varies between 0 (surface) and 100 m (bottom). Those figures illustrate the variation of the scattered and total fields as the depth of the observation point changes. Again the scattered field computed, taking into account multiple interactions with the boundaries, is compared with Ingenito's approximate expression for the scattered field. The most obvious pattern in the differences between the two results is the variation with frequency noted before. The differences decrease as frequency increases. The differences between the full scattered field and Ingenito's approximation are due mostly to the strong bottom reflections originated by the acoustically hard boundary. Other calculations, not exhibited here, show that configuration geometry is also very important. Certain source-target-observation point configurations are more sensitive to multiple scattering effects than others. In particular off-plane scattering, that is, scattering off the vertical plane that contains the source and the scattering object, is more sensitive to those multiple scattering effects than in-plane scattering.
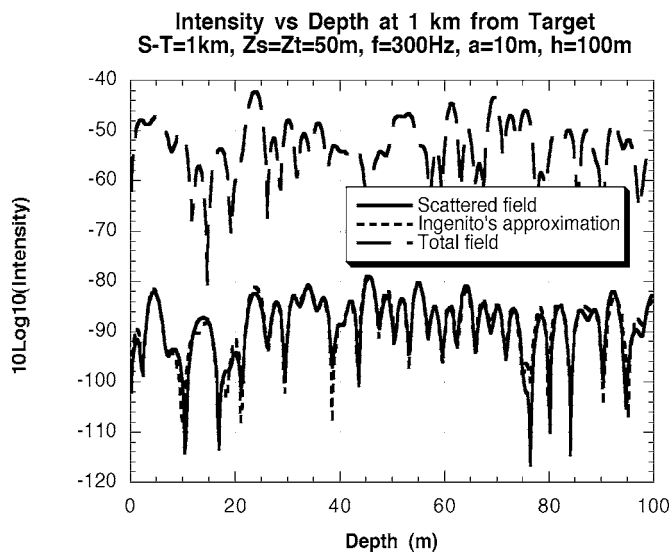
V. SUMMARY

In this paper the theory of acoustic scattering by an object in a bounded medium such as a layered oceanic waveguide was formulated in such a way as to use the scattering amplitude computed in an unbounded, homogeneous medium as input. The main results are an extension of the partial wave expansion method for scattering in an ocean acoustics waveguide and the generalized optical theorem obeyed by the effective scattering amplitudes. The impact of multiple scattering between the object and the boundaries was discussed with both analytical and numerical examples. The circumstances where one expects the greatest impact of those interactions on observable quantities were discussed. The generalized optical theorem for scattering in a bounded environment such as a waveguide was derived, and it shows that the boundary interactions alter the balance between scattered and incident energy in comparison with scattering in an unbounded medium. The scattering formulation presented here is somewhat similar to that of Sammelmann and Hackman.[10] However, here it is in an operational form that is more straightforward to use, for example, this new formulation allowed for a simple and elegant derivation of the optical theorem, which provided significant insights in the physics of scattering in a bounded medium.

ACKNOWLEDGMENT

FIG. 6. This figure is similar to Figs. 4 and 5, except that in this figure the cw point source has a frequency of 300 Hz. This figure illustrates the variation of the scattered and total fields as the depth of the observation point changes.

[1] K. Huang and C. N. Yang, "Quantum-mechanical many-body problem with hard-sphere interaction," Phys. Rev. **105**, 767 (1957).
[2] D. K. Dacol and D. G. Roy, "Wave scattering in waveguides," J. Math. Phys. **44**, 2133 (2003).
[3] D. K. Dacol and D. G. Roy, "Generalized optical theorem for scattering in inhomogeneous media," Phys. Rev. E **72**, 036609 (2005).
[4] F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051 (1987).
[5] A. Sarkissian, "Multiple scattering effects when scattering from a target in a bounded medium," J. Acoust. Soc. Am. **96**, 3137 (1994).
[6] G. C. Gaunard and H. Huang, "Acoustic scattering by a spherical body near a plane boundary," J. Acoust. Soc. Am. **96**, 2526 (1994).
[7] J. A. Fawcett and R. Lim, "Evaluation of the integrals of target/seabed scattering using the method of complex images," J. Acoust. Soc. Am. **114**, 1406 (2003).
[8] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, NY, 1968).
[9] P. G. Bergmann, "The wave equation in a medium with a variable index of refraction," J. Acoust. Soc. Am. **17**, 329 (1946).
[10] G. S. Sammelmann and R. H. Hackman, "Acoustic scattering in a homogeneous waveguide," J. Acoust. Soc. Am. **82**, 324 (1987).
[11] G. C. Gaunard and M. F. McCarthy, "Resonances of elastic scatterers in

fluid half-spaces," IEEE J. Ocean. Eng. **12**, 395 (1987).

[12]Z. Ye and C. Feuillade, "Sound scattering by an air bubble near a plane sea surface," J. Acoust. Soc. Am. **102**, 798 (1997).

[13]P. Ratilal and N. C. Makris, "Extinction theorem for object scattering in a stratified medium," J. Acoust. Soc. Am. **110**, 2924 (2001).

[14]R. G. Newton, *Scattering Theory of Waves and Particles*, 2nd ed.

(Springer-Verlag, New York, 1982).

[15]C. M. Linton, M. McIver, P. McIver, K. Ratcliffe, and J. Zhang, "Trapped modes for off-centre structures in guides," Wave Motion **36**, 67 (2002).

[16]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Melville NY, 1994), Chap. 2, pp. 108–110.

The frequency dependence of compressional wave velocity and attenuation coefficient of intertidal marine sediments

G. B. N. Robb, A. I. Best, J. K. Dix, and J. M. Bull
National Oceanography Centre: Southampton, European Way, Southampton, SO14 3ZH, United Kingdom

T. G. Leighton and P. R. White
Institute of Sound and Vibration Research, University Road, Southampton, SO17 1BJ, United Kingdom

To advance the present understanding of the frequency dependence of compressional wave velocity and attenuation in marine sediments a series of well-constrained *in situ* acoustic transmission experiments (16 to 100 kHz) were performed on intertidal sediments. The processing techniques incorporated *in situ* spreading losses, sediment to transducer coupling and thorough error analyses. Significant variations in velocity and attenuation were observed over scales of tens of meters within the same sediment type. Velocity was generally nondispersive in sands, while highly variable silt velocities prevented any meaningful dispersion estimates from being determined. The attenuation coefficient was proportional to frequency for 75% of the experimental sites. The measured compressional wave properties were compared to predictions from the Grain-Shearing model. For the sandy sites, the phase velocities predicted by the Grain Shearing model exceed those measured, while predicted phase velocities agreed with measured group velocities at specific locations for the silty sites. For both silts and sands predicted dispersions are comparable to the intrinsic errors in group velocity and hence undetectable. The attenuation coefficients predicted by the Grain Shearing model adequately describe the measured attenuation coefficients, within the observed variability.
© 2006 Acoustical Society of America. [DOI: 10.1121/1.2345908]

I. INTRODUCTION

A wide range of commercial and academic marine disciplines, from oil and gas surveying[1] to marine archaeology,[2] require knowledge of the geotechnical properties of the seafloor. The inversion of geoacoustic properties obtained through remote refraction/reflection profiling represents a promising approach for obtaining these *in situ* geotechnical properties for large volumes of seafloor sediment. Conversely, marine sonar applications require accurate predictions of the geoacoustic properties of seabed sediments, in order to enhance object detection through active sonar techniques. In order to perform the inversions/predictions required for these applications, the frequency dependencies of the more commonly used compressional wave properties need to be determined accurately for a range of sediment types. Although there is a long history of good work in this field, the diversity of sediment types and the wide range of frequencies of interest means that further measurements are required to determine the frequency dependencies and to refine theories.

In situ transmission tests offer a compromise between laboratory and remote techniques, and therefore present a promising manner of examining the frequency dependence of compressional wave velocity and attenuation coefficient in marine sediments. A highly flexible experimental geometry allows the compressional wave properties of a well-defined volume of sediment to be measured over a wide frequency range. Furthermore, the sediment disturbance associated with probe deployment can be minimized through careful consideration of the deployment techniques and experimental geometry adopted.

There are, however, other practical difficulties associated with performing such measurements that have tended to limit the number of published *in situ* compressional wave datasets available, generally to a single sediment type or a single frequency. Although it is clear that the use of data from a single research project, using a single acquisition device and common data processing techniques, are inherently more directly comparable, this lack of data has promoted the use of data compilations to incorporate wider ranges of frequencies and sediment types. In order to generate data, which allows generic frequency-dependent relationships to be examined, it is important that the repeatability of the acoustic signal emitted by the source is quantified, a factor that will depend on both the coupling of the transducers to the sediment and the electronic signal transmitted to the source. Some processing techniques also require that corrections for spreading losses are applied, and while the assumption that spreading losses are the same in water and sediment is approximately true in some cases, the use of *in situ* spreading losses would be more relevant.

These experiments utilized a single data acquisition system and enhanced data processing techniques to measure both compressional wave velocity and attenuation coefficient, with corresponding intrinsic errors. The frequency range investigated, i.e., 16 to 100 kHz, spans those used by sub-bottom profilers and those incorporated by laboratory techniques. The frequency dependence of these properties

TABLE I. A summary of published research on velocity dispersion in marine sediments, with **N** denoting no dispersion and velocities expressed in terms of the velocity of water ν_w, where absolute values are not available. A lack of published information is indicated by a dash.

Frequency (kHz)	Velocity	Degree of dispersion	Sediment type	First author and reference
14–100	1704–1712 m s^{-1}	**N**	*In situ* sand	Hamilton (Ref. 5)
15–1500	$1.189 \cdot \nu_w$	**N**	Reconstituted sand/silt/clay	McLeroy (Ref. 6)
0.2–4.7	1410–1740 m s^{-1}	**N**	*In situ* silt	Best (Ref. 7)
375–935	1560–1741 m s^{-1}	**N**	Sand/silt/clay samples	McCann (Ref. 8)
25–100	1727–1797 m s^{-1}	4%	*In situ* sand	Buckingham (Ref. 9)
0.125–400	$1.05 \cdot \nu_w$–$1.17 \cdot \nu_w$	11.4%	*In situ* sand	Williams (Ref. 10)
50–350	-	0.5%	Reconstituted sand	Wingham (Ref. 11)
3–200	$0.93 \cdot \nu_w$–$0.99 \cdot \nu_w$	4%–6%	Artificial silt/clay	Hampton (Ref. 12)
20–100	-	25 m s^{-1}	*In situ* sand	Gorgas (Ref. 13)
1–30	1580–1720 m s^{-1}	8.9%	*In situ* sand	Turgut (Ref. 14)
0.125–50	1580–1755 m s^{-1}	11.1%	*In situ* sand	Stoll (Ref. 15)

were examined using empirical fits, while the validity of the grain-shearing model[3,4] was investigated.

II. PRESENT UNDERSTANDING OF COMPRESSIONAL WAVE PROPERTIES

There is still contention over the frequency dependence of both compressional wave velocity and attenuation coefficient in marine sediments. A significant volume of literature exists to support both nondispersive[5–8] and dispersive[9–15] relationships (see Table I). Doubt has been expressed as to the applicability of laboratory based studies to *in situ* values[6,8,11,12] (owing to sediment disturbance), while the large variations in velocity dispersion observed by different studies (0.5% to 11.4%) highlights the need for a series of experiments that examines a large range of sediment types and frequencies in a well-constrained manner. Also, error analyses presented in earlier pioneering works generally refer to the standard deviation arising from variability. As the field matures, it is necessary to examine instead the intrinsic error associated with the experimental and processing techniques adopted.

The frequency dependence of compressional wave attenuation coefficient α (in dB m^{-1}) is typically expressed using

$$\alpha = k_A f^{\,q}, \tag{1}$$

where k_A is the constant of proportionality[5] and q is the exponent of frequency, f. Published measurements of compressional wave attenuation coefficients in marine sediments are presented in Table II. Though a number of researchers report an attenuation coefficient that is proportional to frequency[5,6,16–20] from 40 Hz to 1.5 MHz, there is some concern over this relationship for a number of reasons. These are namely the deviation from unity of certain published exponents of frequency[5,17,20] or a lack of statistical validation.[16,18,19] The difficulty of reliably determining the frequency dependence of the attenuation coefficient is highlighted by a compilation of data from fine-grained sediments.[21] Although this presents an optimum exponent

of 1.12, it is stated that, within 95% confidence limits, it is impossible to distinguish between this optimum exponent and an exponent of unity. Finally, attenuation coefficients measured at low frequencies[22,23] (5 to 250 Hz) are less than those predicted by the linear extrapolation of higher frequency data,[5] hence implying that the frequency dependence of attenuation coefficient may differ in this low frequency range. Attenuation coefficients that do not vary with frequency in a linear manner are observed by a number of authors.[7,9,12,21,24–27] Alternative data compilations observe "some evidence" for nonlinear frequency dependence, with nonlinear regions occurring between 10 and 100 Hz in sands and 1 to 10 kHz in silts.[23]

III. *IN SITU* EXPERIMENTS

A series of well-constrained *in situ* acoustic transmission experiments were performed on inter-tidal sediments along the south coast of England. These experiments utilized a new acoustic probe system, the Sediment Probing Acoustic Detection Equipment (SPADE).[28] The SPADE consists of a source and a pair of matched receivers (Fig. 1) that were deployed individually in aluminium channels. The use of a fixed rig, which may allow the relative positions of the probes to be more accurately defined, was declined, owing to the associated reduction in the flexibility of the system. The source consists of a section of piezoceramic material with a convex emitting face. This operates in an untuned manner over the usable frequency range of 16 to 100 kHz, while the matched receivers possess sensitivities of −189 to −198 dB *re* 1 V/μPa over this range. The waveforms selected for the transmission experiments consisted of a series of tonal pulses,[28,29] each possessing 10 cycles modulated by a Blackman-Harris envelope[30] with central frequencies from 16 to 100 kHz, in 2 kHz increments. Signals emitted by the SPADE are repeatable to the sampling period of ±1 μs and ±3.2% in amplitude, while water tests and field trails confirmed that received signals accurately resemble shifted, scaled versions of the voltage signals sent to the source transducer.[29]

TABLE II. Summary of published research on the frequency dependence of a compressional wave attenuation coefficient in marine sediments, including measured attenuation coefficients, exponent q of frequency dependence, and sediment type. A lack of published information is indicated by a dash.

Frequency (kHz)	Attenuation coefficient (dB m^{-1})	Exponent q	Sediment type	First author and reference
3.5–100.0	0.6–74.3	0.94–1.1	*In situ* sand to clay	Hamilton (Ref. 5)
15–1500	-	1	Reconstituted silt/clay	McLeroy (Ref. 6)
25–100	8–60	Nonlinear	*In situ* sand	Buckingham (Ref. 9)
0.125–400	1–200	Nonlinear	*In situ* sand	Hampton (Ref. 12)
3.5–100	1.5–55	1	Compilation of data (sands)	Hamilton (Ref. 16)
4–50	0.28–3.00	$1 \pm 15\%$	*In situ* Mud	Wood (Ref. 17)
0.04–0.09	0.59	1	*In situ* clayey silt	Bennett (Ref. 18)
5–50	0.30–1.86	1	*In situ* silty clay	Lewis (Ref. 19)
5–50	-	1.00–1.26	*In situ* sand	McCann (Ref. 20)
0.03-500	5.6×10^{-4}–90	1.12	Compilation of data (muds)	Bowles (Ref. 21)
0.05–1.00	2.2×10^{-4}–9.5×10^{-4}	1.25–1.50	-	Evans (Ref. 24)
20–40	0.10–2.48	0.6–3.4	Sand to clayey silt samples	Shumway (Ref. 25)
100–1000	40–150	1.3–2.0	Silt and clay core samples	Courtney (Ref. 26)
20–300	3–34	Nonlinear	Glass beads	Hovem (Ref. 27)
0.2–4.7	0–4	Nonlinear	*In situ* silt	Best (Ref. 7)

The present project chose to examine intertidal sediments only, as this allowed the positions and orientations of the probes to be measured as accurately, and more easily, than would have been possible for a similar deployment in completely submerged sediments. This, however, introduces the risk of encountering partially saturated sediments. Gas bubbles can be introduced into marine sediment through a variety of mechanisms (e.g., the anaerobic decomposition of organic matter[31]) and have been observed in both intertidal sediments[32] and wholly submerged sediments.[33] Until 2003, bubbles with radii less than 500 μm could not be resolved by the present bubble sizing techniques available, namely, the x-ray CT scanning of pressurized cores.[34] Therefore published *in situ* acoustic measurements from "saturated" sediments have assumed full saturation.

In order to allow the attenuation coefficient of the sediment to be measured, it was essential that the beam patterns associated with the SPADE are known. The receiver's beam pattern[28] varies from an almost omnidirectional field at 1 kHz to a much more directional field at 100 kHz. The relatively simple geometry of the transmission experiments allowed estimated changes in the amplitude of the received signal, which arise from the receiver orientations, to be incorporated into the intrinsic error budget (Sec. IV).

The pressure field emitted by the source was modeled using a modified version of a technique used to predict spreading losses for circular array transducers.[35] This considered the source to be a segment of a cylinder from which sound radiates outward, which was subdivided into a two-dimensional (2-D) grid of elements (Fig. 2). The manipulation of Eq. (10) in Borsboom *et al.*[35] results in the following expression for the pressure at a field-point X and angular frequency ω:

$$P(X,\omega) = W(\omega) \cdot dx \cdot dy \sum_{n=1}^{N_1} \sum_{m=1}^{N_2} \frac{e^{-i\omega R'/v} \cdot e^{-i\omega R_{n,m}/v}}{2\pi R_{n,m}}, \quad (2)$$

where $W(\omega)$ is the spectrum of the selected voltage pulse; dx and dy are the size of the elements in the x and y directions, respectively; N_1 and N_2 are the number of integration steps in each direction; $R_{n,m}$ is the distance from source element n,m to field point X; and R' is the minimum distance between point X and the transducer surface. The pressure response at field-point X associated with a given frequency ω can be
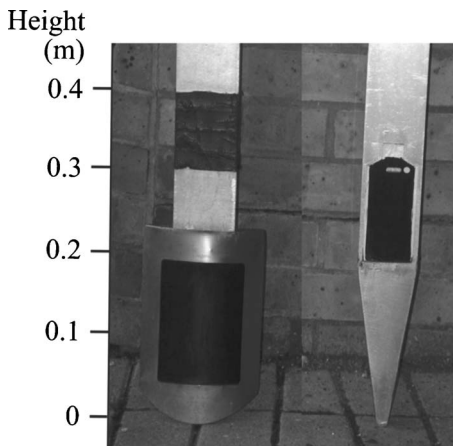
FIG. 1. SPADE source (left) and receiver (right), both inserted in aluminium channels to assist their deployment into the sediment.
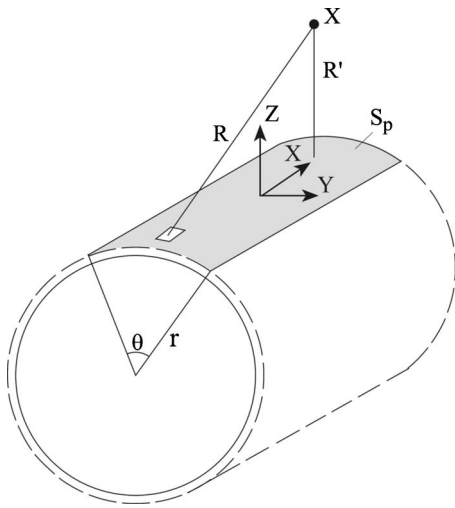
FIG. 2. Geometry used in the simulation of the pressure field emitted by the source, denoted by the shaded area segment S_p. An example field point X is displayed, along with the distance R from source element dS, the minimum distance to the source R', the parameters defining the source (r and θ) and the coordinate system used. From Borsboom *et al.* (Ref. 35).

FIG. 3. SPADE deployment during fieldwork, with image (a) of receivers positioned at approximately the same S-R separation, lying on either side of the central line marked by the measuring pole. Diagrammatic plan view (b) displays measured perpendicular distance, D, and offsets, Y_1 and Y_2, to receivers, with subsequent receiver deployments indicated at positions 2 and 3. In order to clearly display the offsets, the plan view has been rescaled, with offsets used in the fieldwork ranging from 1 to 10% of the S-R separations used.

obtained from the temporal maximum of the inverse Fast-Fourier Transform of $P(X, \omega)$. The pressure field was computed for the horizontal plane that intercepts the center of the source, as the use of a common depth for source and receivers in the fieldwork performed ensured that all receiver deployments lie in this plane. Deviations from this common depth are incorporated into the error budget (Sec. IV).

The model was verified through calibration signals measured using the SPADE source and receivers in a water tank.[29] Transmitted signals were detected by placing the receivers in the horizontal plane that intercepts the center of the SPADE source. The goodness of fit between the observed and modeled decay ranged from 0.79 and 1.00, representing a good to excellent fit. Discrepancies between the predicted and measured decays lay within the error limits in the experimental setup, i.e., $\pm 2°$ in the horizontal angle, $\pm 1°$ in the vertical angle, ± 0.02 m in source-receiver (S-R) separation, and ± 0.01 m in depth. Hence, the pressure model described above adequately predicts the pressure emitted by the SPADE source in nondispersive media.

Pressure fields were simulated for pulses with central frequencies from 16 to 100 kHz (in increments of 2 kHz) and compressional wave velocities from 1300 to 1800 m s^{-1} (in increments of 100 m s^{-1}), i.e., the range of velocity values obtained from field data analysis. Near-to-farfield transitions were computed as the distance from the center of the source to the last axial maximum.[36] These lay less than 0.43 m from the source for all compressional wave velocities measured and all frequencies used.

The experimental procedure involved inserting the source and a pair of receivers vertically into the sediment, with the inclination of the probes to the vertical accurate to $\pm 2°$. The two receivers were deployed at approximately the same S-R separation astride the perpendicular to the source face (Fig. 3). While the experimental/processing techniques adopted herein only require a single source and receiver, the use of two receivers increased the quantity of data that could

be collected. The center of the source and receivers were placed at a common depth of 1 m (± 0.02 m), which reduced the possibility of interfering arrivals from seabed surface reflections, while allowed sediment samples to be readily obtained from the sediment volume through which the acoustic wave propagated. The receivers were positioned at approximately the same S-R separation, with S-R separations calculated from the measured perpendicular distances, D, and offsets, Y_1 and Y_2 [see the plan view in Fig. 3(b)]. S-R separations varied from 1.0–8.1 m, all of which were located in the farfield. A series of acoustics pulses were transmitted, consisting of 5 shots at each central frequency from 16 to 100 kHz (in increments of 2 kHz). The receivers were removed and redeployed at a closer S-R separation before the same series of acoustics pulses was transmitted. The successive use of closer S-R separations ensured that previous deployments did not disturb sediment that later deployments would reexamine. Tidal windows allowed the examination of between three to seven S-R separations at each location.

In the silts, the source and receivers could be deployed through the manual application of a vertical force to the top of the aluminium channels. In the sands this was not possible and hence a modified box corer (width 0.13 m and breadth 0.21 m) was used. This was initially inserted in the sediment to a depth of 1 m and the sand within it excavated. The probes were then inserted and, after the removal of the box corer, the sediment refilled. This allowed the majority of sediment through which the acoustic pulses propagates to remain undisturbed, with the range of disturbed sediment around each probe (<0.1 m) much less than the S-R separations examined.

To minimize the risk of encountering unsaturated sediments, sites were selected to lie in coastal regions of the United Kingdom with relatively low tidal ranges (i.e., less
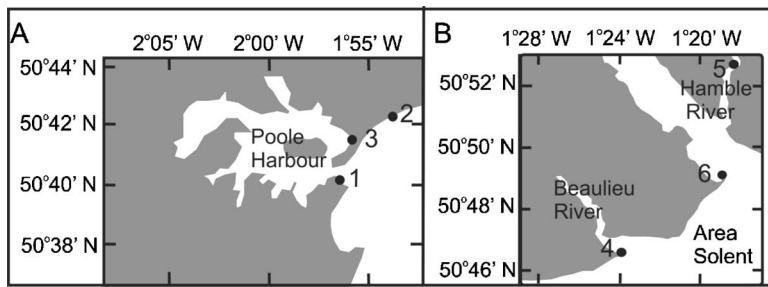
FIG. 4. Geographical location of the intertidal sites examined, including (a) three sandy sites (Sites 1 – 3) and (b) three silty sites (Sites 4 – 6) used in this study. Grey regions denote land.

than 1.5 m). Six sites were examined in total, including three sand sites (Sites 1–3) located at the mouth of Poole Harbour [see Fig. 4(a) and Table III] and three silt sites (Sites 4–6) located in Southampton Water and the Beaulieu estuary [see Fig. 4(b) and Table III]. At each site between two and four specific locations were examined in order to allow any natural sediment variability to be assessed.

At each location sediment samples were collected from a depth of 1 m and the temperature and salinity of the pore water was recorded (see Table IV). The grain size distribution at each site was obtained from between 6 to 12 sediment samples collected from the insonified sediment volume. In addition, a single push core was collected from each site, from which porosity was measured at 1 cm intervals along the core using a multisensor core logger.[37] The geotechnical properties are presented in Table III, which quotes mean values with standard deviations. The sediments were classified using standard sedimentological procedures[38,39] as medium to fine sands, with porosities from 29.3%–34.5%, and coarse to medium silts, with porosities from 63.2%–80.1%. Although great care was taken to select visually uniform sites the actual variabilities in mean grain diameter M varied from 0.07 to 0.35 ϕ for sandy sites and 0.15 to 0.30 ϕ for silty sites. Larger scale heterogeneities, such as shells, shell fragments, and pebbles with diameters from 2 mm to 10 cm,

were observed at Sites 1, 2, and 4.

IV. DATA PROCESSING TECHNIQUES

Established acoustic transmission processing techniques were adapted to both assess the repeatability of the transducer to sediment coupling and to incorporate the sound pressure fields directly relevant to the sediment type under examination. In the present section we summarize these techniques, with the reader referred to Robb et al.[28] for a more detailed discussion. Any deviations from assumptions used by these processing techniques were accounted for through the error analysis employed.

Common preprocessing steps[28] were applied to both the voltage signals sent to the source transducer and the received signals. While the majority of received signals resembled shifted, scaled versions of the voltage signals sent to the source transducer with negligible distortions (Fig. 5), some distorted waveforms were detected. These were attributed to the presence of interfering signals, arising from scatter from sediment heterogeneities and reflections from the sediment surface, and were omitted from the following analysis.

The technique adopted to calculate the group velocity of the sediment optimized data from a range of S-R separations. Arrival times t_A were obtained from a correlation of the ana-

TABLE III. Geotechnical properties of the six intertidal sites examined, including geographical location, sediment type, mean grain diameter M (ϕ), porosity n (%) and clay, sand and silts fractions (%). The ϕ scale (Ref. 39) is defined by $M(\phi)=-\log_2[M(mm)/M_0]$, where M_o represents a "standard" grain diameter of 1 mm.

Site	Location	Sediment type	Mean grain diameter M(ϕ)	Porosity n(%)	Sand (%)	Silt (%)	Clay (%)
1	N 50° 40' 42-47" W 001°56' 56-58"	Mod well sorted Bimodal medium sand	1.58±0.35	29.3±1.0	99.8±0.3	<0.2	<0.2
2	N 50° 42' 23-29" W 001° 54' 29"	Well sorted Unimodal medium sand	1.84±0.15	32.9±2.5	99.0±0.3	<1.0	<1.0
3	N 50° 41' 00-43" W 001°55' 54-55"	Well sorted Unimodal fine sand	2.16±0.07	34.5±1.7	99.6±0.3	<0.4	<0.4
4	N 50° 46' 39" W 001°23' 43-45"	Poorly sorted Unimodal medium silt	6.75±0.30	80.1±2.8	7.7±1.5	81.0±3.3	10.7±3.2
5	N 50° 52' 34-35" W 001° 18' 43-44"	Poorly sorted Unimodal medium silt	6.77±0.20	60.5±2.1	7.6±4.5	82.0±2.9	9.2±3.5
6	N 50° 48' 56" W 001°18' 34"	V. poorly sorted Unimodal coarse silt	5.31±0.15	63.2±0.9	27.5±7.0	69.7±6.5	2.8±0.4

Robb et al.: Frequency-dependence of compressional wave properties

TABLE IV. Pore water properties at sand and silt sites examined, including salinity (S), temperature (T), density, and bulk modulus.

Sites	Salinity $S(0/00)$	Temperature T (°C)	Density (kg m^{-3})	Bulk modulus (GPa)
Sandy (1–3)	0.3–6.3	14–18	999–1003	2.26–2.29
Silty (4–6)	0.3–30.4	7–11	1000–1023	2.09–2.23

lytical received signal with the analytical voltage signal sent to the source transducer. These arrival times are related to the S-R separation d, the group velocity ν and a constant time lag associated with SPADE t_L through

$$t_A = d/\nu + t_L. \tag{3}$$

The group velocity at each frequency was therefore obtained from the reciprocal of the gradient of a linear least-squares fit between arrival time and S-R separation. The total error in arrival time, t_E, was calculated by combining the intrinsic timing error in the digital acquisition card (± 1 μs) with the error introduced through the use of analytical signals at the correlation stage (± 10 μs) in quadrature.[28] This allowed the standard deviation in velocity σ_V, which was adopted as the relevant error, to be computed from

$$\sigma_V = \sqrt{t_E^2 / N(\overline{d^2} - \overline{d}^2)}, \tag{4}$$

where N is the number of data points to which the linear fit is applied and the overscore notation denotes the statistical mean.[40] The intrinsic errors for the dataset examined ranged from ± 20 to ± 70 m s^{-1} in sands and ± 10 to ± 25 m s^{-1} in silts.

The technique developed to calculate attenuation coefficient incorporates both the spreading losses relevant to the sediment under examination and the effects of variable transducer to sediment coupling in the error analysis. The received pulse was resampled to 10 MHz and the signal amplitude $A(f,d)$ assigned the value of the maximum peak-to-peak voltage. The resulting amplitude is a function of the following processes:

$$A(f,d) = V_O(f) \cdot S_T(f) \cdot G(f,d) \cdot R_T(f) \cdot R_e(f) \cdot C \cdot e^{-\alpha_n(f)d}, \tag{5}$$

where spreading losses $G(f,d)$ are a function of frequency f and S-R separation d. The amplitude of the voltage signal transmitted to the source $V_O(f)$, the Transmit Voltage Response of the source transducer $S_T(f)$, the response of the receiving transducers $R_T(f)$, the electronic gain of the receiving amplifiers $R_e(f)$, and the attenuation of the sediment in nepers m^{-1} $\alpha_n(f)$ are all functions of frequency, while the coupling term C is considered constant[28] at each location examined. The comparison of the received signal amplitudes at a single frequency, correction for spreading losses using simulated sound pressure fields applicable to the sediment under examination (see Sec. III), and taking natural-logarithms reduces Eq. (5) to

$$\ln(A_C) = B_1 - \alpha_n d, \tag{6}$$

where A_C represents the amplitudes with spreading losses accounted for and B_1 is a constant. Hence, at each location and frequency, a weighted linear least-squares fit was to applied to $\ln(A_C)$ against d, the gradient of which is equal to the effective attenuation coefficient α_n in nepers m^{-1}.

Intrinsic errors[28] in attenuation coefficient were computed by combining errors arising from the resolution of the acquisition card; the variability of the coupling parameter; the use of pressure fields computed at discrete velocities only; and uncertainties in the measured gain of the receiving amplifiers. For the magnitude of the received amplitudes observed (5.7 mV to 10 V), the error associated with coupling variability represents the dominant element. The approximate error in each value of $\ln(A_C)$, Δ, was obtained from

$$\Delta = A_e / A_C, \tag{7}$$

where A_e is the resulting error in A_C. Hence, the error in the attenuation coefficient, σ_α, was obtained from

$$\sigma_\alpha = \sqrt{\frac{\overline{\overline{\Delta}}}{N(\overline{\overline{d^2}} - \overline{\overline{d}^2})}}, \tag{8}$$

where a double overline denotes a weighted mean.[40] The resulting intrinsic errors ranged from ± 1 to ± 7 dB m^{-1} for all sites examined.

This approach cannot be used without first checking the effect of neglecting the asymmetry of the errors in Δ, an inherent omission arising from the use of symmetric errors in corrected amplitudes. Such analysis of the data showed that the attenuation coefficients obtained using asymmetric errors and those obtained using the approach above differ by less than 0.07 dB m^{-1}, while errors in attenuation coefficient[28] differ by less than 17.5%.
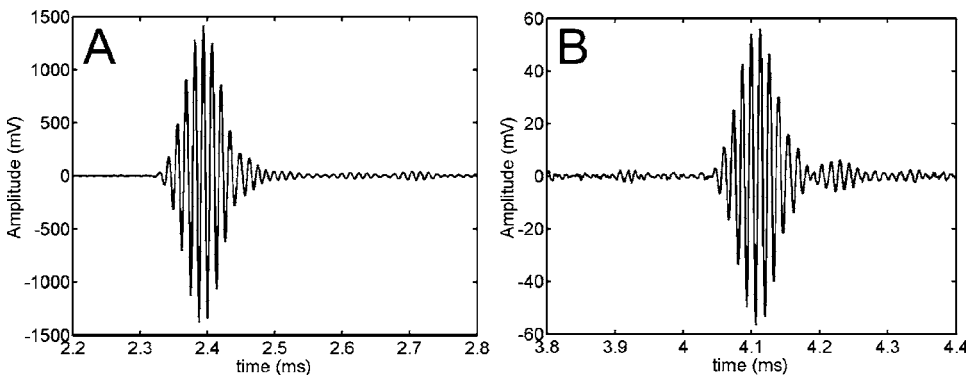
FIG. 5. Example waveforms received at Site 6 (Location 2) with central frequencies of 80 kHz and S-R separations of 2.51 m (a) and 5.00 m (b).
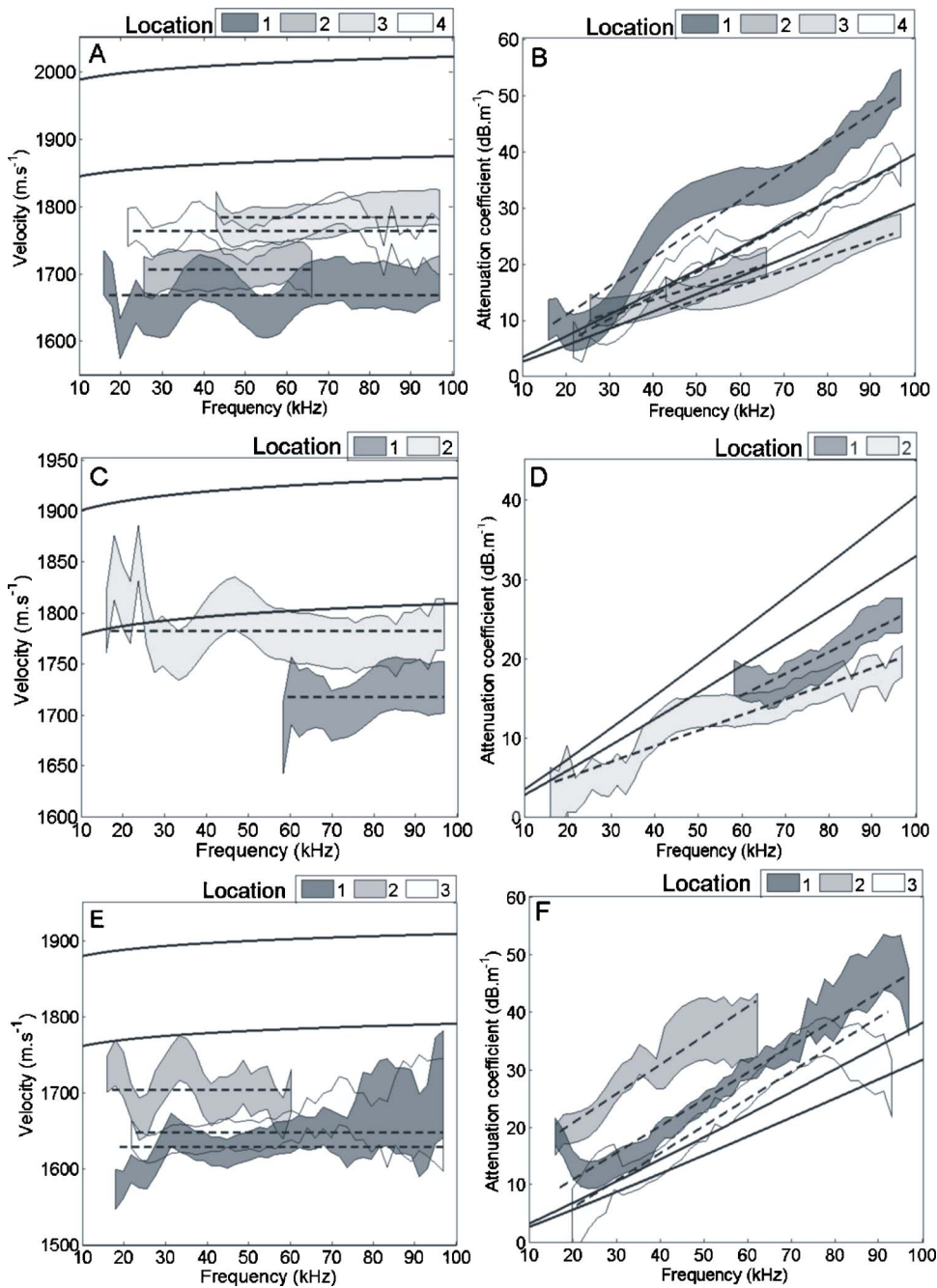
FIG. 6. Compressional wave properties of intertidal sandy sites including velocities for Sites 1 (a), 2 (c), and 3 (e) and attenuation coefficients for Sites 1 (b), 2 (d), and 3 (f). The measured group velocities and attenuation coefficients, with intrinsic errors, are denoted by shaded regions (see legends for specified locations). For each location the weighted mean of velocity and weighted least-squares fit to attenuation coefficient are displayed by dashed lines. The limits of the phase velocities and absorption coefficients predicted by the G-S model are denoted by solid lines.

V. EXPERIMENTAL RESULTS

The compressional wave velocities and attenuation coefficients are displayed in Fig. 6, for predominantly sandy sites (Sites 1 to 3), and Fig. 7, for the predominantly silty sites (Sites 4 to 6). In these figures, data at certain frequencies and locations have been omitted, where distorted waveforms resulted in insufficient degrees of freedom for the application of the linear fits (see Sec. IV).

The common frequency dependencies presented in the literature for velocity and attenuation coefficient were tested for their applicability to these intertidal sites. They include the hypotheses that compressional wave velocity is independent of frequency,[16] and that both compressional wave velocity and attenuation coefficient are proportional to frequency.[5] These hypotheses were tested using a weighted

mean and a weighted linear least-square fit, respectively.[40] The weighted linear least-square fits are expressed as

$$v = v_A \cdot f + c_1, \qquad (9)$$

$$\alpha = k_A \cdot f + c_2, \qquad (10)$$

where v_A, and k_A are the constants of proportionality, and c_1 and c_2 are respective values of velocity and attenuation coefficient corresponding to a frequency of zero. Note that Eq. (10) is a modified version of Eq. (1), with $q=1$. The inclusion of the term c_2 accounts for the possibility of a nonlinear relationship between the attenuation coefficient and frequency outside the frequency range examined here.

For each hypothesis, the deviation of the measured compressional wave property from a weighted fit was converted

Robb *et al.*: Frequency-dependence of compressional wave properties
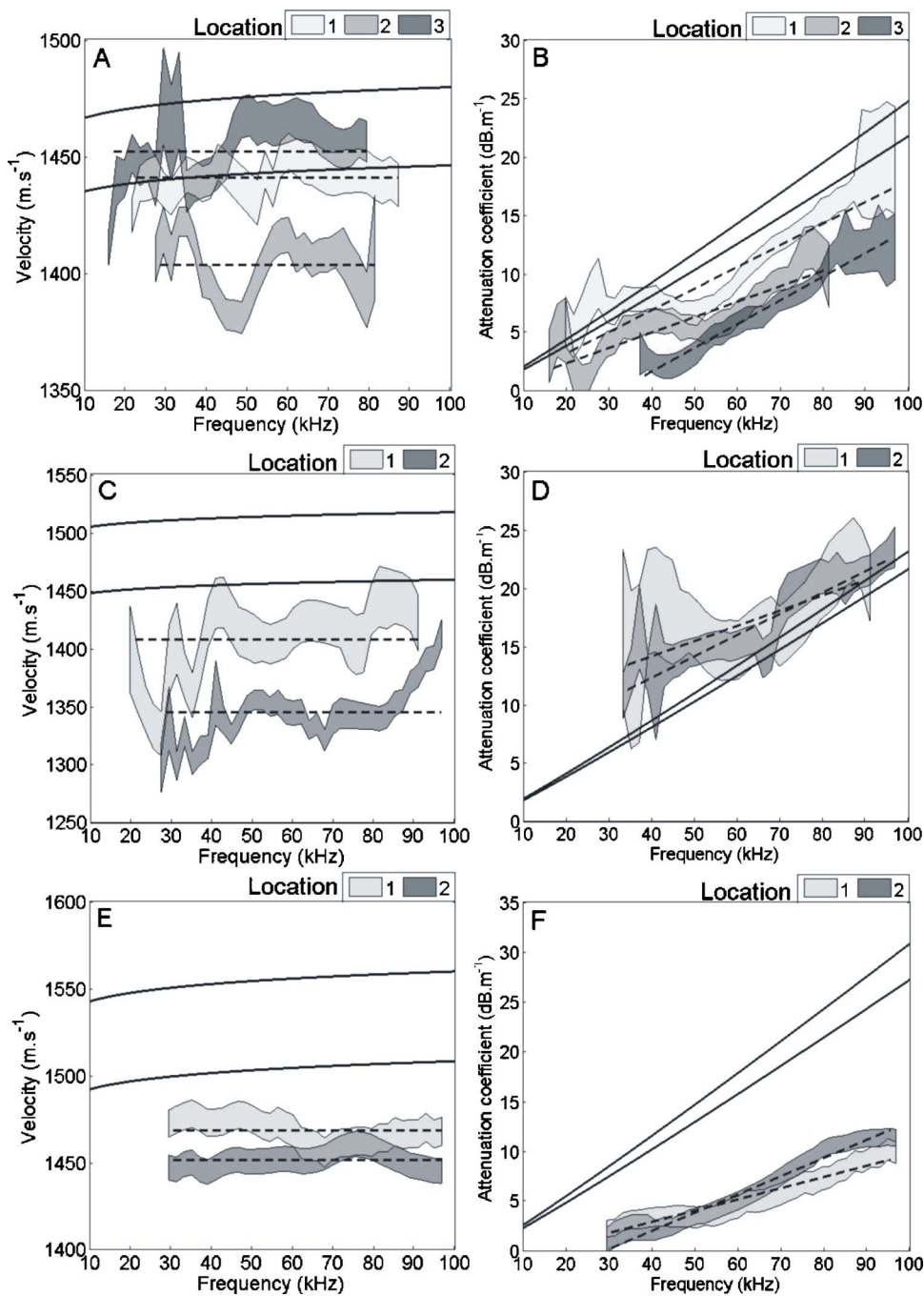
FIG. 7. Compressional wave properties of intertidal silty sites including velocities for Sites 4 (a), 5 (c), and 6 (e) and attenuation coefficients for Sites 4 (b), 5 (d), and 6 (f). The measured group velocities and attenuation coefficients, with intrinsic errors, are denoted by shaded regions (see legends for specified locations). For each location the weighted mean of velocity and weighted least-squares fit to attenuation coefficient are displayed by dashed lines. The limits of the phase velocities and absorption coefficients predicted by the G-S model are denoted by solid lines.

into a confidence limit using the χ^2 distribution.[41] Confidence limits greater than or equal to 95% are considered to indicate an acceptable statistical probability that the weighted fit accurately represents the measured values. The results of the frequency-dependence analysis are displayed in Fig. 6, Fig. 7 and Table V, with standard deviations quoted as the relevant errors. As analysis did not support velocities that were proportional to frequency for either sandy or silty sites (standard deviations in ν_A were greater than ν_A) results from Eq. (9) are omitted from both Table V and the following discussion.

Confidence limits for weighted mean velocities were greater than 95% for seven of the nine predominantly sandy locations. This supports the hypothesis of negligible velocity dispersion from 16 to 100 kHz for these sandy sediments within the observed errors. For the predominantly silty sites, confidence levels in the weighted mean were extremely low for Sites 4 and 5, while nondispersive velocities were supported for Site 6. The velocities at Sites 4 and 5 display a sharply oscillatory nature, particularly at lower frequencies, which conceals any fundamental frequency-dependent trends that may be present. Although the depth at which the probes were deployed (1 m) and the S-R separation used (8.1 m) introduces the possibility of interference from surface reflections for frequencies less than 60 kHz, the omission of distorted waveforms from the analysis (Sec. IV) rules out this interference as a cause of these oscillations. It is more likely that velocities at these sites are dominated by a considerable degree of variability. This argument is supported by the more spatially variable nature (Sites 4 and 5 display relatively

TABLE V. Compressional wave properties of intertidal sites, including the spread of properties at a single frequency; and the results of the frequency-dependent statistical analysis performed (i.e. weighted mean velocity, constant of proportionality for attenuation coefficient k_A and corresponding χ^2 confidence limits).

Site	Velocity spread (m s^{-1})	Atten. spread (dB m^{-1})	Location	Applicability of non dispersive velocity		Applicability of attenuation coefficient proportional to f	
				mean (m s^{-1})	χ^2 Conf. (%)	k_A (dB m^{-1} kHz^{-1})	χ^2 Conf. (%)
1	130	24.4	1	1668±5	>97.5	0.51±0.02	<80.0
			2	1712±6	>99.5	0.14±0.03	>99.5
			3	1783±5	>99.5	0.27±0.03	>99.5
			4	1760±4	<80.0	0.41±0.01	<80.0
2	128	8.7	1	1718±6	>99.5	0.29±0.04	>99.5
			2	1782±4	>97.5	0.19±0.01	>97.5
3	166	18.6	1	1631±4	>95.0	0.52±0.02	>99.5
			2	1703±6	<80.0	0.50±0.05	>99.5
			3	1647±4	>99.5	0.48±0.03	>95.0
4	85	7.3	1	1440±1	>90.0	0.19±0.01	<80.0
			2	1404±1	<80.0	0.13±0.01	<80.0
			3	1455±1	<80.0	0.20±0.01	>99.5
5	109	6.2	1	1408±3	<80.0	0.13±0.04	>99.5
			2	1345±3	<80.0	0.18±0.01	>99.0
6	31	3.1	1	1469±1	>99.0	0.10±0.01	>99.5
			2	1451±1	>99.5	0.17±0.02	>99.5

large variabilities in the mean grain diameter; see Table III) and a more heterogeneous nature (shells were observed at Site 4) of these sediment volumes.

The statistical analysis indicates that, from 16 to 100 kHz, attenuation coefficient is proportional to frequency for both sandy and silty sediments, with confidence limits greater than the threshold value of 95% for three quarters of the locations examined (Table V). Constants of proportionality k_A are greater in sands than in silts, ranging from 0.14 to 0.52 dB m^{-1} kHz^{-1} (with errors less than ±0.05 dB m^{-1} kHz^{-1}) in the sands and 0.1 to 0.2 dB m^{-1} kHz^{-1} (with errors less than or equal to ±0.04 dB m^{-1} kHz^{-1}) in the silts.

For comparison purposes, the velocities were converted to velocity ratios, using pore water velocities (1425 to 1483 m s^{-1}) computed[29] from salinities and temperatures measured at the experimental sites (Table IV). Velocity ratios varied from 1.06 to 1.26 (±1.1 to ±4.5%) at the sandy sites, and 0.92 to 1.04 (errors less than ±1.9%) at the silty sites. While these values broadly agree with velocity ratios published by Richardson *et al.*[42] (0.98 to 1.12 for subtidal medium sands to clays), the velocity ratios for the silts extend below those previously measured. This may be a consequence of the presence of free-gas bubbles in the silts, which may be introduced by either the intertidal nature of the sediments or the biogenic decomposition of organic material (Total Organic Contents ranged from 3% to 11% in the silts). The effect of free gas on the compressional wave velocity depends on the relationship between the insonifying frequencies and the resonant frequencies of the bubbles.[43,44] At frequencies less than resonance, the velocities of partially saturated sediments are less than those of saturated sediments, while at frequencies greater than resonance the velocities of partially saturated and saturated sediments are approximately the same. Of direct relevance to this work are previous *in situ*

measurements of compressional wave velocities in partially-saturated beach sands,[32] which recorded velocity ratios as low as 0.09 for frequencies of 800 Hz. The much higher velocity ratios observed in the present project indicates that either the frequencies examined in this project (16 to 100 kHz) are approximately equal to or greater than the resonance frequencies or the sediments examined in this project possess extremely small fractions of gas bubbles.

Attenuation coefficients ranged from 2 to 52 dB m^{-1} at the sandy sites and 1 to 23 dB m^{-1} at the silty sites, values that agree with previously published results for subtidal sediments over a similar frequency range.[9,21] The variability of geoacoustic properties was examined using the maximum range of velocities and attenuation coefficients spanned by a single frequency (see Table V). The sandy sites possessed the greater variability than the silty sites, particularly for an attenuation coefficient. Unfortunately this could not be related to porosity changes, owing to the collection of a single push core from each site, while no correlation was observed with the observed variability in the mean grain diameter.

VI. MODELING OF *IN SITU* COMPRESSIONAL WAVE PROPERTIES

The measured *in situ* compressional wave properties were compared to those predicted by the grain-shearing (G-S) model.[3,4] The G-S model[3,4] considers marine sediments to act as unconsolidated granular media and considers a new type of dissipation arising from grain-to-grain contacts, which has fundamentally different properties to the more classical dissipation mechanisms of viscosity or Coulomb friction.[45] This intergranular dissipation allows elastic type behavior to arise from intergranular interactions rather than finite frame moduli.

In order to predict the phase speed and attenuation coefficient of the two waves supported (one compressional and one shear), the G-S model requires 14 geotechnical input parameters and frequency.[4] For certain parameters, only generic site-independent values could be determined. A nominal depth of 1 m was used, while the rms grain roughness was evaluated from porosity versus mean grain diameter regression equations presented by Buckingham.[4] The parameters describing the microscopic interactions at grain boundaries, namely the compressional and shear coefficients and strain-hardening index, cannot be measured. These are typically obtained through fitting the G-S model to measured compressional and shear properties at a known porosity, mean grain diameter and depth (which become the reference porosity, mean grain diameter, and depth, respectively). This model fitting step requires the knowledge of at least three of the four acoustic properties predicted by the theory, i.e., compressional and shear phase velocities and attenuation coefficients. As only two parameters have been measured for the intertidal sediments under consideration, this step cannot be undertaken. Hence the compressional and shear coefficient, strain-hardening index and reference porosity, mean grain diameter and depth that are used in the present implementation of the G-S model are taken from previously published implementations of the G-S model, values that can be considered to be fixed for all siliciclastic sediments.[4]

For the remaining parameters, a range of input values could be determined for each site examined and therefore the limits of predicted velocities and attenuations could be predicted for each site. *In situ* measured values were used for porosities and mean grain diameters (see Table III), while the properties of the fluid, namely bulk modulus and density, were determined from measured salinities and temperatures using the standard equations of state[46] (see Table IV). The densities and bulk moduli of the grains were obtained from the literature,[29] with densities ranging from 2650 to 2750 kg m^{-3} and bulk moduli ranging from 32 to 49 GPa.

The compressional wave properties predicted by the G-S model are displayed in Figs. 6 and 7 for the sandy and silty sites, respectively. For the sandy sites the phase velocities predicted exceeded those measured. Predicted dispersions were less than 1.7% ($<$35 m s^{-1}), both of which are comparable to the intrinsic errors in group velocity and are therefore undetectable. Measured attenuation coefficients were adequately described, within the observed variability, by the predicted attenuation coefficients (2.6 to 40.5 dB m^{-1}).

It is interesting to note that the compressional wave velocities predicted by Wood's Equation[47] for a suspension agrees more closely with measured velocities than those predicted by the G-S model (for the relevant range of physical parameters compressional wave velocities predicted by Wood's Equation range from 1700 to 1821 m s^{-1} for Site 1, 1629 to 1780 m s^{-1} for Site 2, and 1619 to 1738 m s^{-1} for Site 3). This indicates that the values presented for the compressional and shear coefficients for subtidal sands[4] overpredict normal and tangential stresses associated with intergranular sliding that are present in the intertidal sediments examined in the present project. The difference in these stresses between subtidal and intertidal sediments can be

postulated to be due to the different processes that govern the supply, deposition, and dynamics of the sediment in each environment.

For the silty sites the predicted phase velocities (1435–1559 m s^{-1}) agree well with measured velocities at Site 4 and tend to overestimate measured group velocities at Sites 5 and 6. Dispersions are 1.1% (17 m s^{-1}), which are either less than or comparable to the intrinsic errors in group velocity ($<\pm25$ m s^{-1}) and hence undetectable. Attenuation coefficients predicted by the G-S model vary from 1.8 to 30.8 dB m^{-1} for the frequency range examined. These broadly agree with measured attenuation coefficients at Sites 4 and 5 and overestimate measured attenuation coefficients at Site 6.

While the agreement between measured and predicted attenuations implies that the loss mechanism incorporated by the G-S theory is applicable to both sandy and silty sediments, it is interesting to consider the effects of additional scattering losses. It is unlikely that Rayleigh scattering from sediment grains is a significant source of energy loss, as the smallest wavelength examined (16 mm) was much greater than the maximum grain diameter encountered ($<$1 mm). This is supported by the observation that Rayleigh scattering is only significant for frequencies greater than 500 kHz in sands[10] with mean grain diameters of approximately 414 μm (1.27ϕ), i.e., coarser than those examined within this project. Any additional scattering losses present are more likely to arise from larger scale scattering centers or spatial variability in the sediment across the insonified volumes (see Sec. III).

VII. CONCLUSIONS

In this work we focused on ascertaining the frequency dependence of compressional wave velocity and an attenuation coefficient in marine sediments over the frequency range 16 to 100 kHz. This was achieved through a series of well-constrained *in situ* transmission experiments performed in intertidal sediments. The processing techniques adopted incorporated *in situ* spreading losses, a quantitative assessment of coupling and waveform variability, and a thorough error analysis.

The compressional wave velocity ratios ranged from 1.06 to 1.26 (±1.1 to $\pm4.5\%$) at the sandy sites and 0.92 to 1.04 ($<\pm1.9\%$) at the silty sites. Attenuation coefficients ranged from 2 to 52 dB m^{-1} at the sandy sites and 1 to 23 dB m^{-1}, with intrinsic errors less than ±7 dB m^{-1}. A significant variability was observed in both the velocity and attenuation coefficient between the measurement locations, which are separated by tens of meters in geologically homogenous sediments. This suggests that acoustic propagation measurements made over distances on the meter-to-decimeter scale could be difficult to translate to the average seabed acoustic properties of a larger area. The sandy sites possessed greater variability than the silty sites, particularly in the case of an attenuation coefficient.

An analysis of the common frequency dependencies observed in the literature indicated that, from 16 to 100 kHz, velocity was nondispersive for the majority of the sandy locations examined. For the silty locations, the frequency de-

pendency of velocity was more difficult to ascertain, with nondispersive velocities only supported at Site 6. This is attributed to the more spatially variable and heterogeneous nature of the silty sites concealing any fundamental frequency-dependent trends present. The attenuation coefficient was found to be proportional to frequency from 16 to 100 kHz for both sand and silt sediments, within confidence limits of 95%, for three-quarters of the locations examined.

The grain-shearing (G-S) model was applied as detailed in Buckingham.[4] For the sandy sites the phase velocities predicted by the G-S model exceed those measured. This indicates that the stresses associated with intergranular sliding differ between the subtidal and intertidal environment, a possible consequence of the different processes that govern the supply, deposition, and dynamics of the sediment in each environment. For the silty sites, the predicted phase velocities agree with measured group velocities at specific locations. For both silts and sands, predicted dispersions are comparable to the intrinsic errors in group velocity and hence undetectable. The attenuation coefficients predicted by the G-S adequately describe the measured attenuation coefficients, within the observed variability. Finally, it should be noted that the validation of the geoacoustic models examined has been considerably impaired by the variability of the measured velocities and attenuation coefficients.

[1] R. E. Sheriff and L. P. Geldart, *Exploration Seismology* (Cambridge University Press, Cambridge, 1995).

[2] J. M. Bull, R. Quinn, and J. K. Dix, "Reflection coefficient calculation from marine high resolution (Chirp) data and application to an archaeological case study," Mar. Geophys. Res., **20**, 1–11 (1998).

[3] M. J. Buckingham, "Wave propagation, stress relaxation and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am. **108**, 2798–2815 (2000).

[4] M. J. Buckingham, "Compressional and shear wave properties of marine sediments: Comparisons between theory and data," J. Acoust. Soc. Am. **117**, 137–152 (2005).

[5] E. L. Hamilton, "Compressional-wave attenuation in marine sediments," Geophysics **37**, 620–646 (1972).

[6] E. G. McLeroy and A. DeLoach, "Sound speed and attenuation, from 15 to 1500 kHz, measured in natural seafloor sediments," J. Acoust. Soc. Am., 1148–1150 (1968).

[7] A. I. Best, Q. J. Huggett, and A. J. K. Harris, "Comparison of *in situ* and laboratory acoustic measurements on Lough Hyne marine sediments," J. Acoust. Soc. Am., **110**, 695–709 (2001).

[8] C. McCann, "An investigation of the acoustical properties of natural materials," Ph.D. thesis, Department of Physical Oceanography, University College of North Wales, 1967.

[9] M. J. Buckingham and M. D. Richardson, "On tone-burst measurements of sound speed and attenuation in sandy marine sediments," IEEE J. Ocean. Eng. **27**, 429–453 (2002).

[10] K. L. Williams, D. P. Jackson, E. I. Thoros, D. Tang, and S. Schock, "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous material," IEEE J. Ocean. Eng. **27**, 413–428 (2002).

[11] D. J. Wingham, "The dispersion of sound in sediment," J. Acoust. Soc. Am. **78**, 1757–1760 (1985).

[12] L. D. Hampton, "Acoustic properties of sediments," J. Acoust. Soc. Am. **42**, 882–890 (1967).

[13] T. J. Gorgas, R. H. Wilkens, S. S. Fu, L. N. Frazer, M. D. Richardson, K. B. Briggs, and H. Lee, "In situ acoustic and laboratory ultrasonic sound speed and attenuation measured in hererogeneous soft seabed sediments: Eel River shelf, California," Mar. Geol. **182**, 103–119 (2002).

[14] A. Turgut and T. Yamamoto, "Measurements of acoustic wave velocities and attenuation in marine sediments," J. Acoust. Soc. Am. **87**, 2376–2383 (1990).

[15] R. D. Stoll, "Velocity dispersion in water-saturated granular sediment," J. Acoust. Soc. Am., **111**, 785–793 (2001).

[16] E. L. Hamilton, "Acoustic properties of sediments," in *Acoustics and the Ocean Bottom*, edited by A. Lara-Saenz, C. Ranz-Guerra, and C. Carbo-Fite (Cosejo Superior de Investigaciones Cientificas, Madrid, 1987), pp. 4–58.

[17] A. B. Wood and D. E. Weston, "The propagation of sound in mud," Acustica **14**, 156–162 (1964).

[18] L. C. Bennett, "*In situ* measurements of acoustic absorption in unconsolidated sediments," Trans., Am. Geophys. Union, **48**, 144 (1967) (abstract only).

[19] L. F. Lewis, "An investigation of ocean sediments using the deep ocean sediment probe," Ph.D. thesis, Department of Ocean Engineering, University of Rhodes Island, 1971.

[20] D. M. McCann and C. McCann, "The attenuation of compressional waves in marine sediments," Geophysics **34**, 882–892 (1969).

[21] F. A. Bowles, "Observations on attenuations and shear wave velocity in fine-grained marine sediments," J. Acoust. Soc. Am. **101**, 3385–3397 (2000).

[22] R. D. Stoll, "Marine sediment acoustics," J. Acoust. Soc. Am., **77**, 1789–1799 (1985).

[23] A. C. Kibblewhite, "Attenuation of sound in marine sediments: A review with emphasis on new low frequency data," J. Acoust. Soc. Am., **86**, 716–738 (1989).

[24] R. B. Evans and W. M. Carey, "Frequency dependence of sediment attenuation in two low-frequency shallow-water acoustic experimental datasets," IEEE J. Ocean. Eng., **23**, 439–447 (1992).

[25] G. Shumway, "Sound speeds and absorption studies of marine sediment by a resonance method," Geophysics **25**, 451–467 (1960).

[26] R. C. Courtney and L. A. Mayer, "Acoustical properties of fine-grained sediments from Emerald Basin: Toward the inversion for physical properties using Biot-Stoll model," J. Acoust. Soc. Am. **93**, 1145–1154 (1993).

[27] J. M. Hovem and G. D. Ingram, "Viscous attenuation of sound in saturated sand," J. Acoust. Soc. Am. **66**, 1807–1812 (1979).

[28] G. B. N. Robb, A. I. Best, J. K. Dix, P. R White, T. G. Leighton, J. M. Bull, and A. Harris, "The measurement of the *in situ* compressional wave properties of marine sediments," IEEE J. Ocean. Eng., in press.

[29] G. B. N. Robb, "The *in situ* compressional wave properties of marine sediments," Ph.D. thesis, University of Southampton, Southampton, 2004.

[30] S. G. Schock, L. R. LeBlanc, and L. A. Mayer, "Chirp sub-bottom profiler for quantitative sediment analysis," Geophysics **54**, 445–450 (1989).

[31] D. D. Rice and G. E. Claypool, "Generation, accumulation, and resource potential of biogenic gas," AAPG Bull. **65**, 5–25 (1981).

[32] R. Bachrach and A. Nur, "High-resolution shallow-seismic experiments in sand, Part 1: Water table, fluid flow and saturation," Geophysics **63**, 1225–1233 (1998).

[33] P. Fleischer, T. H. Orsi, M. D. Richardson, and A. L. Anderson, "Distribution of free gas in marine sediments: a global overview," Geo-Mar. Lett. **21**, 103–122 (2001).

[34] A. H. Reed and K. B. Briggs, "Gas bubbles in marine mud-How small are they?," J. Acoust. Soc. Am. **114**, 2318 (abstract only) (2003).

[35] J. M. G. Borsboom, E. I. Cespedes, A. F. W. Van der Steer, C. T. Lancee, and E. F. Deprettere, "Simulation of circular array ultrasound transducers for intravascular applications," J. Acoust. Soc. Am. **108**, 827–835 (2000).

[36] LE. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals in Acoustics* (Wiley, New York, 1982).

[37] A. I. Best and D. G. Gunn, "Calibration of multi-sensor core logger measurements for marine sediment acoustic impedance studies," Mar. Geol. **160**, 137–146 (1999).

[38] G. M. Friedman and J. E. Sanders, *Principles in Sedimentology* (Wiley, New York, 1978).

[39] M. R. Leeder, *Sedimentology: Process and Product* (George Allen and Unwin, London, 1982).

[40] R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (Wiley, Chichester, 1989).

[41] C. F. Dietrich, *Uncertainty, Calibration and Probability* (Adam Hilger, Bristol, 1991).

[42] M. D. Richardson, D. L. Lavoie, and K. B. Briggs, "Geoacoustic and physical properties of carbonate sediments of the Lower Florida Keys," Geo-Mar. Lett. **17**, 316–324 (1997).

[43] A. L. Anderson and L. D. Hampton, "Acoustics of gas bearing sediments I, Background," J. Acoust. Soc. Am. **67**, 1865–1889 (1980).

[44] A. L. Anderson and L. D. Hampton, "Acoustics of gas bearing sediments II. Measurements and models," J. Acoust. Soc. Am. **67**, 1890–1903 (1980).

[45] M. J. Buckingham, "Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments," J. Acoust. Soc. Am. **102**, 2579–2596 (1997).

[46] P. Siedler, "Properties of seawater," in *Numerical data and functional relationships in science and technology*, edited by J. Sunderman (Springer-Verlag, Berlin, 1986), pp. 237–259.

[47] A. B. Wood, *A Textbook of Sound* (G. Bell and Sons, London, 1941).

Sound speed and attenuation measurements in unconsolidated glass-bead sediments saturated with viscous pore fluids

Brian T. Hefner[a] and Kevin L. Williams

Applied Physics Laboratory, College of Ocean and Fisheries Sciences, University of Washington, 1013 N.E. 40th Street, Seattle, Washington 98105

As part of a recent ocean sediment acoustics experiment, a number of independent sound speed and attenuation measurements were made in a well-characterized sandy sediment. These measurements covered a broad frequency range and were used to test both Biot-Stoll theory and Buckingham's more recent grain-to-grain shearing model. While Biot theory was able to model the sound speed well, it was unable to predict the attenuation measured above 50 kHz. This paper presents a series of measurements made in the laboratory on a simple glass-bead sediment. One goal of these measurements was to test the hypothesis that the attenuation measured at-sea was a result of scattering from shells within the sediment. The laboratory sediments used were saturated with fluids with different viscosities in order (assuming that Biot-Stoll theory is correct) to shift the dispersion into the frequency range of the measurement system. The measured attenuation in the glass-bead sediments exhibited the same frequency dependence as observed in the ocean experiment even though no shells were present. The laboratory results motivated development of a sediment model which incorporates both fluid viscosity and grain-to-grain interactions as embodied in a simple frequency-dependent, imaginary frame modulus first suggested by Biot. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354030]

I. INTRODUCTION

In the study of sediment acoustics, there is an ongoing debate about the nature of dispersion and attenuation in unconsolidated, sandy sediments. Many measurements have been made in sandy sediments, both *in situ* and in the laboratory, as well as in simpler unconsolidated glass-bead sediments. The results of these experiments have led to two differing conclusions. Some measurements seem to indicate that attenuation in these sediments varies as the first power of the frequency, f^1, with little or no significant dispersion. This view has found its most vocal support in Hamilton[1] and has recently been revisited by Buckingham who argues that this conclusion is supported by current research into the properties of dry sand.[2–4] Both Hamilton and Buckingham have argued that the sediment is best described by a viscoelastic model. Both of these approaches determine the relevant parameters through empirical fits to the data limiting their predictive value.[5] The approaches differ, however, in that Hamilton's approach is purely empirical while Buckingham believes that the attenuation is due to friction between the sand grains.[4] Buckingham postulates that the relevant parameters can be determined from the fluid and sand properties although it is not clear at this point in the theory's development how this would be done.

Due to the porous nature of the sediment, some researchers theorize that the relative motion of the pore fluid and the sediment frame should lead to viscous damping of the acoustic wave. Biot theory provides a successful description of sound propagation through porous materials[6,7] and has thus been applied to modeling sand sediments. The attenuation predicted by Biot theory, however, varies as f^2 at low frequencies and $f^{1/2}$ at high frequencies which contradicts some of the measurements in sandy sediments. The sound speed is also predicted to exhibit significant dispersion in the frequency range where the transition between these two limiting attenuations takes place. Some researchers, most notably Stoll,[8] note that there are deviations from the f^1 frequency dependence in many measurements as well as evidence of dispersion indicating that viscous losses due to the motion of the pore fluid may be important. Indeed, many measurements have been successfully modeled with Biot theory by assuming complex frame moduli to account for frictional losses at the grain contacts. This modified theory has come to be known as Biot-Stoll theory.[9] Biot theory has an advantage over the viscoelastic models of Hamilton and Buckingham in that the relevant parameters in the theory can in principle be measured, although not without some effort. Biot-Stoll theory remains empirically driven, however, in that the complex frame moduli are generally determined from a best fit to the data.

With the number of measurements available, it might seem strange that a consensus has not been reached on the proper description of sound propagation in sediments. The problem lies not with a lack of experimental effort but rather in the difficulties inherent in performing these measurements. Usually it is not possible to make sound speed and attenuation measurements over a frequency range broad enough to observe the dispersion predicted by Biot-Stoll theory nor is it usually possible to fully characterize the sedi-

---

[a]Electronic mail: hefner@apl.washington.edu
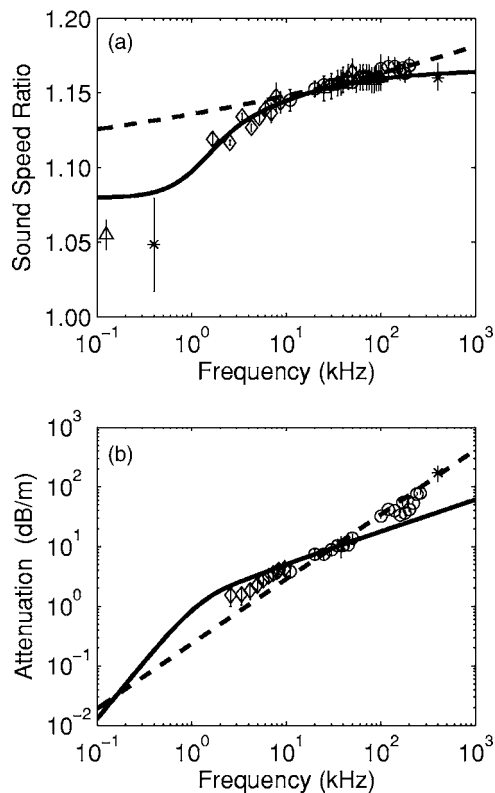
FIG. 1. (a) Sound speed and (b) attenuation measured during SAX99. Details of these measurements can be found in Ref. 12. Also shown are the sound speed and attenuation predictions for Biot theory (solid line) and the Buckingham model (dashed line) using the parameters given in Table I.

ment. Realizing the importance of overcoming these difficulties, a number of researchers, utilizing different measurement techniques, came together to determine both the sound speed and attenuation in the sediment at a single site during a recent high-frequency sediment acoustics experiment (SAX99).[10,11] Extensive efforts were made to determine the properties of the sediment and several researchers measured the sound speed in the sediment over a frequency range of 125 Hz–400 kHz and attenuation over a frequency range of 2.6–400 kHz. Details of these measurements can be found in Ref. 12.

Despite the broad range of frequencies and careful measurement of the sediment properties, the results of the SAX99 experiments have not been successful in resolving the sediment acoustics debate. Both Biot-Stoll theory and Buckingham's model have been applied to the data and both have their respective strengths and weaknesses as shown in Fig. 1. Biot-Stoll theory captures the dispersion in the sound speed across most of the frequencies with the exception of the two lowest frequency measurements. The measured attenuation deviates, however, from that predicted by Biot-Stoll theory at the highest frequencies and appears to follow the f^1 frequency dependence observed previously. Consequently, Buckingham's model fits the attenuation well at these high frequencies with only a slight deviation at the lower frequency measurements. His theory is not as successful in modeling the sound speed as it predicts a curvature in the dispersion that is opposite that of the measured values and only fits the data over a limited frequency range.

As Williams et al.[12] points out, these results are one of four separate measurements that indicate that the sound speed in sand sediments at low frequencies is lower than at high frequencies. This is consistent with the predictions of Biot and the authors speculate that the deviation in attenuation at high frequencies may be due to scattering within the sediment from either volume inhomogeneities or shell and shell fragments. In order to test this hypothesis, a series of laboratory measurements were made in simple glass-bead sediments. The measurements were performed using one of the systems which was used during SAX99 to measure sound speed and attenuation over the 100–260 kHz frequency range. Although this range was extended down to 60 kHz for the laboratory experiments, it is still not low enough to observe significant dispersion for glass beads saturated with water. To overcome this limitation, a second glass-bead sediment was saturated with silicone oil, which has a viscosity of approximately 0.1 Pa s. Biot theory predicts that by increasing the viscosity, the region of significant dispersion should shift upward in frequency and hence should be observable with the measurement system. Biot theory also predicts that attenuation scales with viscosity. Details of the sediments as well as the measurement system are discussed in Sec. II. This experiment is similar to the measurements made by Seifert et al.[13] In those measurements, several different fluids were used to vary the viscosity of the pore fluid from 0.001 to 1 Pa s. However, they observed an f^2 frequency dependence in the attenuation whereas the measurements presented here show an f^1 dependence over a similar range of frequencies. For the measurements in a silicone oil saturated sediment, significant dispersion is observed in the current experiments. These results are similar to those of SAX99 in that the attenuation deviates from Biot theory over the same frequency range while the sound speed is well described by Biot theory. The measurement results and model comparisons are shown in Sec. III. These results do not support the scattering hypothesis for the SAX99 results. While Biot does not capture all aspects of the data, a model which combines the effects of viscosity (Biot theory) with frame losses (perhaps attributable to grain-to-grain shearing[4] or squirt flow[14]) is able to describe all three data sets. This model is developed in Sec. IV while the implications of the data-model comparisons using this hybrid model are discussed in Sec. V.

II. EXPERIMENTAL PROCEDURE

The system used to make the laboratory sound speed and attenuation measurements was previously used during at-sea experiments.[10,12] This diver-deployed system is able to make sound speed measurements in the 100–200 kHz frequency range and attenuation measurements in the 100–260 kHz frequency range. The system consists of four ITC 6148 transducers (two receivers and two transmitters) which allow measurements to be made over four separate paths. During the experiment the path lengths were from 22 to 44 cm and the transducers were inserted into the sediment to a depth of 10 cm. Because this system was designed to be diver-deployed, it was relatively compact and thus easy to use for laboratory based measurements.

TABLE I. Parameters used in calculating the effective density fluid model and Buckingham model results. The hybrid model results were also calculated using these parameters except for the material exponent and rigidity coefficient.

Parameter	Value		
	SAX99	Water/beads	Si100/beads
Bulk modulus of grains or glass beads (K_g)	3.2×10^{10} Pa	7.0×10^{10} Pa	7.0×10^{10} Pa
Mass density of grains or glass beads (ρ_g)	2690 kg/m^3	2500 kg/m^3	2500 kg/m^3
Bulk modulus of pore fluid (K_f)	2.395×10^9 Pa	2.23×10^9 Pa	9.68×10^8 Pa
Mass density of pore fluid (ρ_f)	1023 kg/m^3	1000 kg/m^3	968 kg/m^3
Sound speed pore fluid (c_f)	1530 m/s	1493 m/s	1000 m/s
Viscosity of pore fluid (ν)	0.001 05 kg/m s	0.001 kg/m s	0.098 kg/m s
Porosity (β)	0.385	0.384	0.36
Permeability (κ)	2.5×10^{-11} m^2	1.1×10^{-10} m^2	8.43×10^{-11} m^2
Tortuosity (α)	1.35	1.45	1.47
Material exponent ($n=m$)	0.0901	0.0467	0.3902
Ragidity coefficient ($\gamma_T = \gamma_P + (4/3)\gamma_S$)	2.6823×10^8 Pa	3.871×10^8	1.472×10^6 Pa

The sediments in the laboratory measurements were composed of well-sorted glass-beads which had diameters in the range of 0.43–0.30 mm and a minimum roundness of 90%, according to the manufacturer's specifications. Two fluids were used to saturate the glass beads: water and a silicone oil with a viscosity of 0.098±0.0048 Pa s as determined by the manufacturer. Biot theory predicts that by increasing the viscosity of the pore fluid, the dispersion in the sound speed should shift to higher frequencies.[7] By shifting the dispersion up in frequency, it should be observable in the frequency range of the attenuation array. It should also provide a second test of Biot theory in that the attenuation should scale with the viscosity of the pore fluid. In the measurements of Seifert *et al.* this scaling was not observed leading them to the reject Biot's porous medium description of the unconsolidated sand sediment.[13]

In order to ensure that bubbles would not be present in the sediments, a vacuum was applied to the fluid before the sediment was added and allowed to degas over several days. In order to withstand the vacuum, two stainless steel oil drums were chosen as tanks for the sediments. These drums had an inner diameter of 54.6 cm and a depth of 83.5 cm. The glass beads were added to the degassed fluids in 10 cm layers and the glass bead/fluid combination was degassed and then stirred after each layer was added. Over the course of the experiments, the sediments were stirred and a vacuum was applied to the tanks periodically to remove any bubbles that may have formed. The glass-bead/water sediment had a depth of 37 cm with a water layer above it that was 36 cm thick. The glass-bead/silicone oil sediment had a depth of 23 cm with a 22-cm-thick layer of silicone oil above it.

The properties of each sediment, including that of SAX99, are given in Table I. The properties of the SAX99 sediment were measured using a number of different methods the details of which are given in Ref. 12. The bulk modulus and density of the glass beads were provided by the manufacturer, Potters Industries Inc., and are consistent with reported values of glass found in the literature. The bulk modulus of the water was determined from the average sound speed measurement in the water. The density and vis-

cosity of the silicone oil was provided by the manufacturer, Silchem, and the bulk modulus of the oil was determined from its measured sound speed.

The remaining properties important for the application of Biot theory and which are often difficult to measure in sand sediments are the porosity, β, permeability, κ, and the tortuosity, α. The porosity of the glass-bead/water sediment was determined by taking samples of the sediment with a known volume and measuring the density before and after removing the water by drying the sample in an oven. The value of 0.384 is reasonably close to the random close packing limit for identical spheres. The same procedure could not be applied to the silicone oil sediment because the vapor pressure of the oil is much higher than that of water and a sample of the sediment could not be dried. The random close packed limit of 0.36 was assumed here instead. It may seem reasonable to assume that the porosity should be the same in both cases, but this assumption was rejected because the macroscopic properties of the sediments were found to be significantly different. The silicone oil sediment was found to heal much faster than the water sediment when stirred or when transducers were inserted and removed from the sediment. Also, during degassing, bubbles were able to move out from the interior of the silicone oil sediment much faster and easier than from the water saturated sediment. It is hypothesized that this is due to the higher viscosity of the silicone oil which allows the beads to slide relative to one another during settling with less friction due to hydrodynamic lubrication.[15] It seems more likely that the beads are able to reach a denser ground state in silicone oil than they would in water. (Note: if we assume that the porosity of the silicone oil/glass bead mixture was equal to that of the water/glass bead mixture, the dispersion predicted by Biot theory would still shift to the same frequencies, while the mean sound speed would be lower. This dependence of the mean sound speed on the porosity can be seen in the sensitivity analysis shown in Fig. 1 of Ref. 12. The conclusions drawn from the data/model comparisons in Sec. III therefore remain unchanged.)

The hydraulic permeability was calculated using the Kozeny-Carmen relation,[16]

$$\kappa = \frac{\beta^3 d^2}{180(1-\beta)^2}, \qquad (1)$$

where d is the average diameter of the glass beads. This relationship was tested for the water/glass-bead sediment using a constant head permeameter and the value calculated using the Kozeny-Carmen relation was found to fall within the error of the measurement. A measurement of the tortuosity is notoriously difficult to make for unconsolidated sediment and as a result an expression derived by Berryman for a collection of spheres was used.[17] The expression,

$$\alpha = \frac{2}{1+\beta}, \qquad (2)$$

relates the tortuosity to the porosity and gives a lower value than the expression that was used previously by Berryman to describe sintered glass-bead packs.[18] As Berryman shows in Table 1 of Ref. 17, the expression given in Eq. (2) fits the analytical values for ordered unconsolidated glass-bead packs whereas the previous expression is a better fit to sintered glass-bead packs. As a result, Eq. (2) is used for the tortuosities of the glass-bead sediments. The permeability and tortuosity were measured for the SAX99 sediment and the values used for the data/model comparisons were chosen such that they provided the best fit to the sound speed data while still falling within the uncertainty for the measurement.[12] The use of these values for the SAX99 sediment as opposed to results of Eqs. (1) and (2), reflects the availablity of measurements of the material parameters of SAX99 sediment and the fact that the above mentioned equations are only applicable to collections of spheres.

Because the diameter of the tanks were smaller than the original attenuation array, the array was reconfigured for use in the tanks. The new arrangement maintained the original electronics and transducer mounts, while repositioning the transducers such that they sat at the corners of a rectangle. For each transmitter, one receiver was mounted close to provide a short path length, while the second transducer was mounted further away to provide a longer path length. The sound speed was determined using the difference in phases between a pulse received at two different transducers.[19] As a result, the difference in path lengths, ΔP, is the important distance for this measurement. For the water/glass-bead sediment $\Delta P_1 = 25.99$ cm and $\Delta P_2 = 25.95$ cm while for the silicone oil/glass-bead sediment $\Delta P_1 = 14.39$ cm and $\Delta P_2 = 14.97$ cm. The reduction in path length for the silicone oil was necessary in order to overcome the signal loss due to the increased attenuation. The attenuation was determined from the spectral amplitudes using the method described in Ref. 19. This technique also depends on the difference in the two path lengths. The sound speed and attenuation measurements for both sediments were performed using the original attenuation array over a frequency range of 80–260 kHz. However, due to the large attenuation, measurements above 200 kHz could not be made in the silicone oil saturated sediment.

For each measurement, the transducers were first placed in the overlying fluid layer and the sound speed and spectral amplitude of the pulse in the fluid were measured. The transducers were then inserted into the sediment. The depth of the tranducers in the water/glass-bead sediment was 10 cm while in the silicone oil/glass-bead sediment it was 11.4 cm. The tank was then shaken by hand for approximately 2 min then allowed to sit for 4–6 h to allow the sediment to settle. The sound speed and spectral amplitude were then measured and after the transducers were removed from the sediment, the sound speed and spectral amplitudes in the fluid were again measured. Once the transducers were removed from the tank, the sediment was again shaken and allowed to settle overnight.

In order to extend the lower limit of the frequency range of the array, a Reson TC4040 was introduced which was able to operate in the 60–100 kHz range. The difference in path length for the new array was $\Delta P_1 = 7.82$ cm in water and $\Delta P_1 = 4.53$ cm in silicone oil. Unfortunately, there was an unresolved problem with the phase of the output pulse of the new array which produced an offset in the measured sound speed. Consequently these phase speed measurements were discarded. The attenuation measurements were unaffected by this problem and are presented here.

III. RESULTS AND MODEL-DATA COMPARISONS

The sound speed ratio and attenuation measured in each sediment are presented in Fig. 2. Each data point is the mean of measurements of approximately 20 realizations of each sediment, where a realization represents a complete cycle of the shaking procedure discussed earlier. The error bars on each data point reflect the uncertainties in the path lengths as well the distribution of the measured values. As predicted by Biot theory, the sound speed in the silicone oil/glass-bead sediment exhibits greater dispersion than the water/glass-bead sediment. However, contrary to Biot theory, both media show a frequency dependence for the attenuation that is greater than $f^{1/2}$. A linear fit to the attenuation shows that the attenuation follows an $f^{1.06}$ dependence for the water/glass-bead sediment and an $f^{0.99}$ dependence for the silicone oil/glass-bead sediment.

Note that in Fig. 2(a), the size of the data circles represent the uncertainty associated with the precision of the measurements (without the uncertainty in the position of the transducers). This indicates that there is a peak in the sound speed centered at 160 kHz. This apparent curvature is most likely a consequence of errors due to the response of the transducers at the edges of the operational frequency range of the attenuation array and not a property of the medium.

In order to test Biot theory against these data, it is necessary to know certain material parameters in order to determine the sound speed and attenuation. A majority of these parameters are discussed in the previous section and are given in Table I, but there are two notable exceptions. For both glass-bead sediments, the frame bulk modulus, K_b, and the shear modulus, μ_b, are unknown. Williams has shown that in most sand sediments, because the bulk and shear frame moduli are much smaller than the other moduli, they
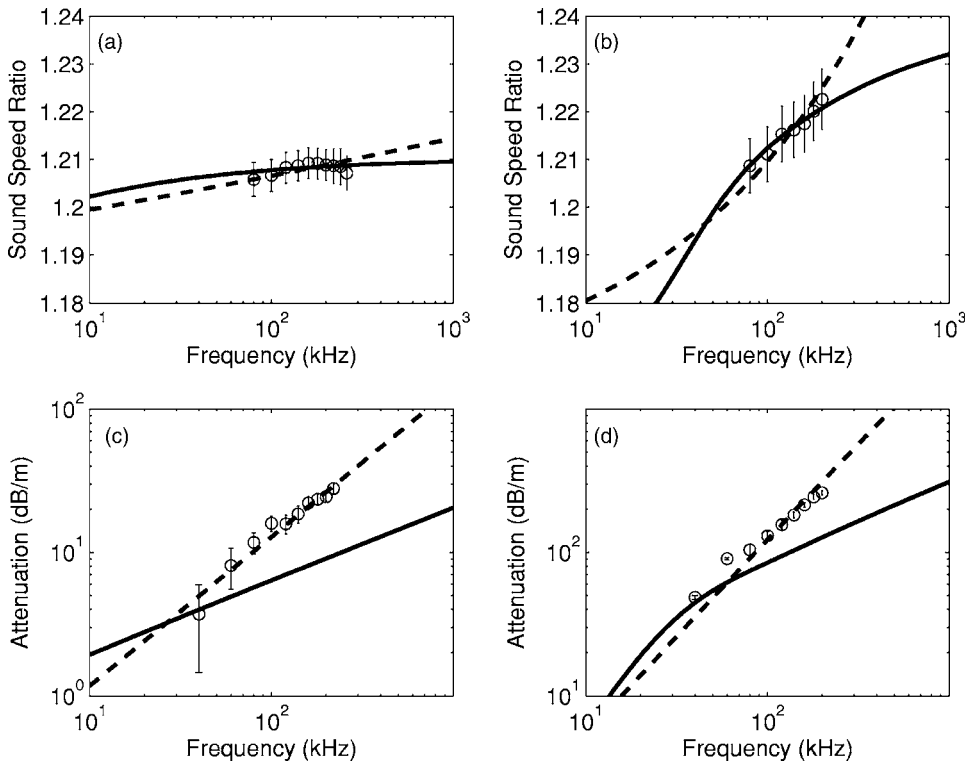
FIG. 2. Sound speed measurements made in a glass-bead sediment saturated with (a) water and (b) silicone oil. Attenuation measurements made in a glass-bead sediment saturated with (c) water and (d) silicone oil. Also shown are the predictions of the EDFM (solid line) and the Buckingham model (dashed line) using the parameters given in Table I. The error bars on the data points represent uncertainties in the position of the hydrophones as well as the distribution of multiple measurements in each sediment. The size of the data circles in (a) represent the uncertainty in the precision of the measurements (without the uncertainty in the position of the transducers).

can be neglected with only minor errors (0.5%) in the high frequency sound speed and attenuation.[20] Taking the $K_b = \mu_b = 0$ limit of Biot theory yields the effective density fluid model (EDFM) which gives the complex wave number for the compressional wave. The EDFM along with the fluid, solid, and sediment properties in Table I yield the solid lines in Fig. 2. In both sediments, the EDFM captures the measured sound speed values but does not capture the high frequency behavior of the attenuation. Although the EDFM does not capture the details of the attenuation, it does predict an increase in attenuation between the water and silicone oil sediments due to the increase in viscosity that is on the same order as the increase in the measured attenuation. Consequently, both the frequency scaling and frequency dependence contradict what was measured by Seifert et al.[13] Note that in the experiments by Seifert et al., the sediment samples were compressed in a triaxial cell while in the current measurements the sediment was allowed to settle under gravity. This compaction could potentially influence the acoustical behavior of the medium and lead to attenuation values different from those observed in the present study. While the results presented here contradict the results of the experiments by Seifert, the current results are consistent with those of Costley and Bedford.[21] In their work, they varied the viscosity from 0.001 to 0.003 Pa s and found that the attenuation at 105 kHz followed the predictions of Biot theory over this range. In those experiments, the measurements were performed in a tank, similar to the work presented here, as opposed to the traxial cell used by Seifert.

The compressional wave number in the grain-to-grain shearing model of Buckingham can be expressed as[4]

$$k^2 = \frac{\omega^2 \rho}{H + \gamma_T(-i\omega T)^n}, \quad (3)$$

where ω is the radial frequency, $\rho = (1-\beta)\rho_s + \beta\rho_f$ is the average density of the sediment, $H = (\beta/K_f + (1-\beta)/K_s)^{-1}$ is the bulk modulus of the sediment evaluated from Wood's equation, $\gamma_T = \gamma_p + 4/3\gamma_s$ is the rigidity coefficient, and n is the material exponent. Note that the rigidity coefficient is a combination of Buckingham's compressional and shear rigidity coefficients. Since a shear wave measurement was not made and cannot be used to determine the values of n and γ_s, the rigidity coefficients can be combined into a single constant which must be fit to the data along with the material exponent. Note that the assumption of $n = m$ made in Buckingham's derivation has also been made here.

The sound speed and attenuation given by Buckingham's model are shown as the dashed lines in Fig. 2 and the values of γ_T and n are given in Table I. For the water/glass-bead sediment, the values of these parameters are the same order of magnitude as those determined for the SAX99 data. For the silicone oil sediment, n is an order of magnitude larger while γ_T is two orders of magnitude smaller. It may be that the large viscosity of the silicone oil brings into doubt the approximation made in Eqs. (58) and (59) in Ref. 4, which stems from the assumption that viscous effects at the grain contacts are negligible. If these approximations are not made, the number of unknowns increases to three (this again assumes that the shear and compressional values are the same), the determination of which is beyond the scope of this paper.

IV. COMBINING FLUID VISCOSITY AND FRAME LOSS EFFECTS

The success of Biot theory in *predicting* (a) the frequency dependence of the dispersion and (b) the scaling of the attenuation with viscosity indicates that the relative motion of the pore fluid and the sediment frame plays a significant role in sound propagation through fluid-saturated granular materials. Any theory of sound propagation should therefore account for this interaction. The failure of Biot theory to account for the frequency dependence of the attenuation suggests that a second loss mechanism is present. Although Buckingham's model was able to fit the attenuation data, the model does not have any predictive ability at this point in its development. However, it does suggest that grain-to-grain interactions may be the second mechanism that is missing from the Biot description of the media. As a starting point, a model is developed here that combines two attenuation mechanisms: viscous flow and grain-to-grain shearing. Buckingham, in his model derivation, examines the need to account for viscous flow losses and concludes that they are negligible compared to grain friction losses. This combined model is again addressed here, but instead of assuming a simple wave propagation model, the grain-to-grain shearing losses are incorporated directly into Biot theory.

In Ref. 4, Buckingham derives expressions for the mean stress in a unit volume of granular material due to a large number of grain-to-grain contacts. The stress associated with radial sliding due to grains being compressed normal to one another is

$$\sigma_{zz}(t) = \lambda_p h_p(t) \otimes \frac{dv_z(t)}{dz} \tag{4}$$

and associated with translational sliding due to shearing at the grain contacts is

$$\sigma_{xz}(t) = \eta_s h_s(t) \otimes \frac{dv_x(t)}{dz}, \tag{5}$$

where λ_p and η_s are constants which contain information about the microsliding events, $h_{p,s}(t)$ are the material impulse response functions (MIRFs) for radial and translational shearing, and the symbol $\otimes$ denotes a temporal convolution (for a detailed discussion of radial and translational sliding see Ref. 4).

Instead of considering the grains and fluid as a single composite medium as in Ref. 4, the grains and fluid will be treated as separate constituents as in Biot theory. Although the grains are not locked together to form a matrix as in Biot theory, the grain-to-grain interactions give rise to a viscoelastic matrix through the "pores" of which the fluid is allowed to move. The elastic constants of this matrix can be determined by considering the components of the strain tensor and how they relate to the stresses given by Eqs. (4) and (5). Consider first the shear strain in the frame matrix,

$$\epsilon_{ij} = \frac{1}{2}\gamma_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{6}$$

This strain will produce a shear stress within the matrix due to the translational sliding of the grains against on another. From Eq. (5), the shear stress is

$$\sigma_{ij} = \eta_s h_s(t) \otimes \left\{\frac{dv_i(t)}{dx_j} + \frac{dv_j(t)}{dx_i}\right\}, \tag{7}$$

where $v_i = \dot{u}_i$. Assuming that the disturbance in the matrix has a harmonic time dependence, $e^{-i\omega t}$, then Eq. (7) becomes

$$\sigma_{ij} = 2\mu_b(\omega)\epsilon_{ij}, \tag{8}$$

where

$$\mu_b(\omega) = -i\omega\eta_s H_s^*(i\omega) \tag{9}$$

and $H_s^*(i\omega)$ is the complex conjugate of the temporal Fourier transform of $h_s(t)$.

For the second elastic modulus of the granular matrix, consider the response of a unit cube of the matrix to a uniform pressure. This pressure gives rise to a change in the volume of the cube equal to $e = \nabla \cdot \mathbf{u}$ and will also produce radial sliding at the grain contacts due to the compression of the grains. Using Eq. (4), the stress-strain relation in this situation becomes

$$\sigma = \lambda_p \nabla \cdot [h_p(t) \otimes \mathbf{v}]. \tag{10}$$

Once again assuming a harmonic compression, the stress-strain relation becomes

$$\sigma = K_b(\omega)e, \tag{11}$$

where

$$K_b(\omega) = -i\omega\lambda_p H_p^*(i\omega) \tag{12}$$

and $H_p^*(i\omega)$ is the complex conjugate of the temporal Fourier transform of $h_p(t)$. The dynamic elastic moduli of the matrix are therefore the grain-to-grain shear modulus, $\mu_b(\omega)$, and the grain-to-grain bulk modulus, $K_b(\omega)$.

The temporal Fourier transforms of the MIRFS in Eqs. (9) and (12) are given approximately by[4]

$$H_s^*(i\omega) \approx \frac{\Gamma(1-m)}{(-i\omega t_s)^{1-m}} \tag{13}$$

and

$$H_p^*(i\omega) \approx \frac{\Gamma(1-n)}{(-i\omega t_p)^{1-n}}, \tag{14}$$

where Γ is the complement of the incomplete gamma function, m, n are the material exponents, and $t_{s,p}$ are the stress relaxation time constants. The approximations are valid assuming that $\omega t_{s,p} \ll 1$. With these expressions, the elastic moduli can be written as

$$\mu_b(\omega) = \gamma_s(-i\omega T)^m \tag{15}$$

and

$$K_b(\omega) = \gamma_p(-i\omega T)^n, \tag{16}$$

where the shear and compressional rigidity coefficients are

$$\gamma_s = \frac{\eta_s}{t_s} \Gamma(1-m)\left(\frac{t_s}{T}\right)^m \tag{17}$$

and

$$\gamma_p = \frac{\lambda_p}{t_p} \Gamma(1-n)\left(\frac{t_p}{T}\right)^n. \tag{18}$$

The time constant T has been introduced to keep terms raised to the fractional powers n and m dimensionless. The expressions are independent of T and hence it is taken to be $T = 1$ s. These expressions for the frame moduli can be used in Biot's equations. A detailed discussion of Biot's equations can be found in Ref. 9 while a concise presentation is found in Eqs. (1)–(8) of Ref. 12.

The form of the expressions for the moduli given by Eqs. (15) and (16) are not new and were first suggested by Biot in the development of a phenomenological approach to solid dissipation in porous medium[6,7] and later explored in detail, outside the context of Biot theory, by Kjartansson.[22] Kjartansson was motivated by early measurements on absorption of sound in rocks which found that the energy lost per cycle or wavelength, Q, was independent of the frequency. To account for this, he proposed a model for the elastic moduli which has essentially the same form as Eqs. (15) and (16). For the shear modulus, and likewise for the bulk modulus, the phase angle between the stress and the strain can be determined by rewriting the modulus as

$$\mu_b(\omega) = \gamma_s |\omega T|^m \exp\left(-i\frac{m\pi}{2}\mathrm{sgn}(\omega)\right). \tag{19}$$

The Q is directly related to this phase angle and from this expression for the shear modulus,

$$\frac{1}{Q} = \tan\left(\frac{m\pi}{2}\right). \tag{20}$$

Because of this property, the model is often referred to as the constant Q (CQ) model.

Kjartansson showed that this model could be expressed as an infinite set of viscoelastic Zener elements.[22] As a consequence of this construction, the CQ model is causal. Buckingham has also shown this directly in the development of his model.[2] The introduction of CQ frame moduli into Biot theory was investigated by Turgut,[23] who numerically integrated the Kramers-Krönig relations and also found that this viscoelastic Biot theory was causal. The models proposed by Biot and Kjartansson and explored by Turgut were all phenomenological and did not make any assumptions about the physics behind the model.

To examine the properties of the hybrid model [the Biot model incorporating frame loss via Eqs. (15) and (16)], it is easiest to look first at the behavior of the shear wave. In the hybrid model, the shear wave number becomes

$$k_s^2 = \omega \frac{\rho_{\mathrm{eff}}^{(s)}}{\mu_b(\omega)}, \tag{21}$$

where

$$\rho_{\mathrm{eff}}^{(s)} = \rho_f \left(\frac{\alpha(1-\beta)\rho_s + \beta(\alpha-1)\rho_f + \dfrac{i\rho\beta\eta F(\omega)}{\rho_f \kappa \omega}}{\rho_f(\alpha-\beta) + \dfrac{i\beta\eta F(\omega)}{\kappa\omega}} \right) \tag{22}$$

and μ_b is the viscoelastic shear modulus given by Eq. (15). For a detailed derivation of this expression see p. 19 of Ref. 9. All of the constants in $\rho_{\mathrm{eff}}^{(s)}$ can be measured directly. Following Buckingham, the material exponent, m, and the rigidity coefficient, γ_s, can be evaluated given a measurement of the shear speed and attenuation at a single frequency. Writing the shear modulus as

$$\gamma_s(-i\omega T)^m = \gamma_s(\omega T)^m(\cos(m\pi/2) - i\sin(m\pi/2)), \tag{23}$$

the material exponent can then be found from

$$m = \frac{2}{\pi} \tan^{-1}\left(\frac{\mathrm{Im}(\rho_{\mathrm{eff}}^{(s)}/k_s^2)}{\mathrm{Re}(\rho_{\mathrm{eff}}^{(s)}/k_s^2)}\right), \tag{24}$$

and the rigidity coefficient is then

$$\gamma_s = \frac{\omega^2 \rho_{\mathrm{eff}}^{(s)}}{(-i\omega T)^m k_s^2}, \tag{25}$$

where $k_s = \omega/c_s - i\alpha_s$, c_s is the shear wave speed, and α_s is the attenuation. During SAX99, the shear wave speed and attenuation were measured[11] at 1 kHz and were $c_s = 120$ m/s with a range of $97-147$ m/s and $\alpha_s = 30$ dB/m with a range of $21-40$ dB/m. Using these values, the material exponent, m, has a best fit of 0.0506 to the mean sound speed and attenuation. The rigidity coefficient has a best fit of 1.82×10^7 Pa to the mean values of the sound speed and attenuation.

The shear speed and attenuation are shown as the solid lines in Fig. 3. Also shown in this figure are the predictions of the Buckingham model and the predictions of Biot theory with a complex shear modulus with no frequency dependence. For this complex shear modulus, the best fit to the mean values of the sound speed and attenuation is $2.83 \times 10^7 - i2.25 \times 10^6$ Pa.

In Ref. 2, Buckingham attempts to address the question of whether it is necessary to account for viscous losses due to global or local motion of the pore fluid when modeling sound propagation in marine sediments. In order to do this, he introduces viscous loss terms into the equation of motion for the saturated sediment. The solution to the resulting wave equation indicates a frequency dependence for both the sound speed and attenuation from which he is able to draw two conclusions. In his combined model, the intergranular friction appears to dominate at lower frequencies while the pore fluid viscosity dominates at high frequencies. Also, the sound speed varies with frequency as $f^{1/2}$ in the high frequency regime where the pore fluid viscosity dominates.

The asymptotics of Eq. (21) indicate that in both high and low frequency limits, the intergranular friction is the dominant loss mechanism, while from Fig. 3 both mechanisms play important roles in the midfrequency range. This is similar to the results of Biot-Stoll theory. Brunson has shown that Biot theory with complex shear moduli is able to account for details of the measured attenuation, specifically the
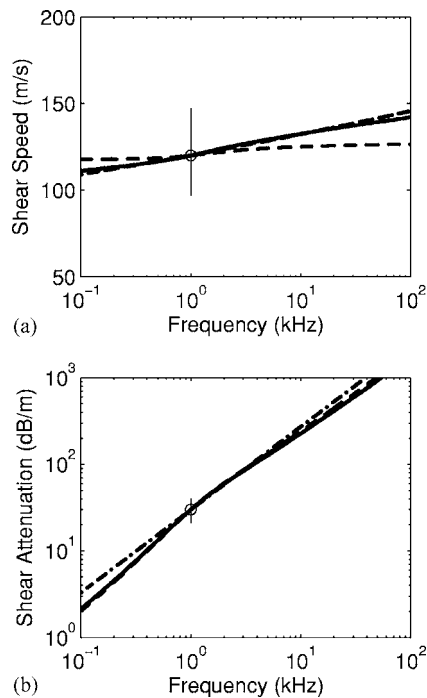
FIG. 3. (a) Shear wave speed and (b) attenuation. The data point at 1 kHz is a shear wave measurement made during SAX99 by Richardson (Ref. 11). The shear wave speed and attenuation predicted by Biot theory (dashed line), the Buckingham model (dot-dashed line), and the hybrid model (solid line). Each of these models was fit to the data point.

"knee" in Fig. 6 of Ref. 24 that a linear fit would miss. (See also Fig. 1 of Ref. 3 for further deviations from the f^1 dependence for shear attenuation.) From the similarities in the behavior of the hybrid model and Biot-Stoll theory in Fig. 3 it seems that the hybrid model would also be able to capture these details in the attenuation. A true test, as seen in Fig. 3(a), would be a measurement of the shear speed across a large frequency range, however this measurement may be prohibitively difficult in sandy sediments.

In Buckingham's model, the grain-to-grain friction is due to a number of random micro-asperity sliding events, each one of which is assumed to be identical regardless of whether the grain-to-grain motion is radial or compressional. Since the material exponents characterize these individual events, it is reasonable to assume that they have the same values. Buckingham presents limited evidence in the form of data-model comparisons that support this conclusion for his model. However, assuming that $n=m$ in the hybrid model does not yield a curve which will fit the at-sea or laboratory attenuation. In fact, the resulting attenuation is nearly indistinguishable from the Biot-Stoll result. This suggests that m and n may not be equal. For the remainder of this analysis we will relax this constraint and allow n to be a free parameter. Using the values for the shear rigidity coefficient and the corresponding material exponent, the remaining coefficient and exponent can be determined for the compressional wave through a best fit to the sound speed and attenuation data. The data-model comparisons presented here support the conclusion that n and m are not equal.

The compressional sound speed and attenuation predicted by the hybrid model for the SAX99 sediment are pre-sented in Figs. 4(a) and 4(d). Both n and γ_p were determined from a best fit to the data and were found to be $n=1.088$ and $\gamma_p=41.96$ Pa. This hybrid model succeeds in combining the effects of fluid viscosity and frame loss and follows the data better than either the Biot or Buckingham model individually. The model also contradicts Buckingham's conclusions about the importance of losses due to viscous flow for the compressional wave. Whereas he concludes that viscous loss is the dominant mechanism at high frequencies, the hybrid model predicts that viscous loss dominates in the midfrequencies while granular friction is the dominant mechanism at high and low frequencies. This is similar to the asymptotics of the shear wave where granular friction also dominates at both the high and low frequencies. Since the frequency dependence of the frame bulk modulus is nearly linear, for low frequencies granular friction will dominate the attenuation since the attenuation due to viscous flow goes as f^2. Also, since Biot theory gives an $f^{1/2}$ dependence at high frequencies, granular friction will again dominate asymptotically.

In order to compare the hybrid model to the data from the glass-bead sediments, four variables need to be determined: n, m, γ_p, and γ_s. Unlike the at-sea data there are no shear wave measurements to use to determine the shear rigidity coefficient and exponent. The results for the at-sea data, however, suggest that the shear modulus may be negligible compared to the other moduli. Setting $\mu_b=0$ for the hybrid model yields both sound speed and attenuation predictions that are nearly indistinguishable from the EDFM predictions. Consequently, for the glass-bead sediments, the shear rigidity coefficient is taken as $\gamma_s=0$. The predictions of the hybrid model are shown in Figs. 4(b) and 4(e) for the water and glass-bead sediment and in Figs. 4(c) and 4(f) for the silicone oil and glass-bead sediments.

For water and glass beads, the grain-to-grain shearing parameters were determined to be $n=0.915$ and $\gamma_p=182.94$ Pa. For the silicone oil and glass beads, the grain-to-grain shearing parameters were determined to be $n=1.059$ and $\gamma_p=57.89$ Pa. The hybrid model captures the data very well in both cases and with values for the grain-to-grain parameters that are consistent with the at-sea sediment. This differs from the Buckingham model for which the material exponent increases by a factor of 10 relative to the water/glass-bead sediment. This change in the material exponent indicates that for the Buckingham model the increase in attenuation and dispersion must be due to a *decrease* in the strain-hardening at the micro-asperities even though the viscosity of the pore fluid was increased by a factor of 100. In the hybrid model, the increase in attenuation is due to the increase in the viscosity of the pore fluid, while the attenuation due to the grain-to-grain friction is nearly identical to that observed in the previous glass-bead sediment and in the SAX99 sediment.

Not only are the grain-to-grain shearing parameters similar in all three sediments, but the material exponent is very nearly equal to one, as can be seen in Table II. This suggests that the hybrid model with $n=1$ may yield predictions that compare favorably to the data. The results of the hybrid model with $n=1$ and γ_p taken as the free parameter
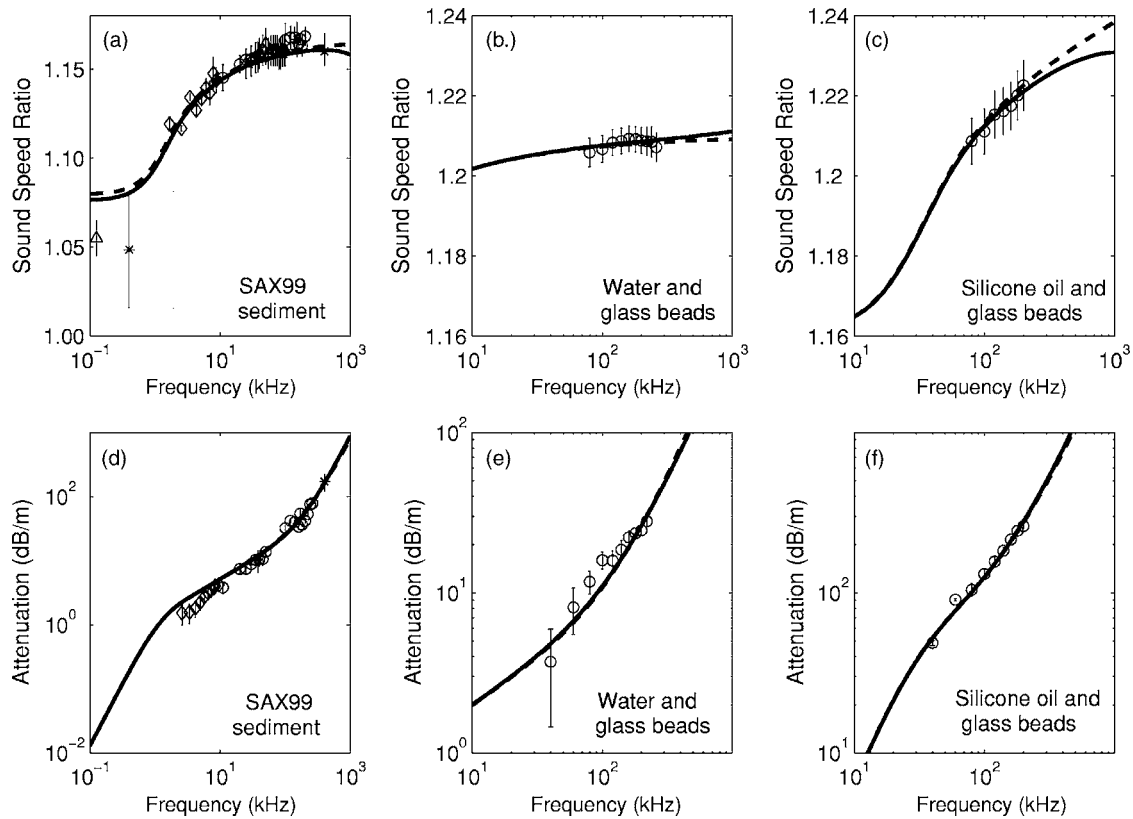
FIG. 4. (a)–(c) Sound speed and (d)–(f) attenuation predictions using the hybrid model for the sediment at SAX99, the water and glass-bead sediment, and the silicone oil and glass-bead sediment. The grain-to-grain shearing parameters in the hybrid model were determined in two ways. For the solid line, the values of both the compressional rigidity coefficient, γ_p, and the material exponent, n, which gave the best fit to both the sound speed and attenuation data were found. For the dashed line, the material exponent was set to $n=1$ while the value of γ_p was determined from the best fit.

are shown as the dashed lines in Fig. 4. For the attenuation, the two curves are practically indistinguishable while for the sound speeds the deviation is most pronounced at high frequencies where the grain-to-grain losses are greatest. Even with this deviation, the model still compares favorably to the data.

V. HYBRID MODEL WITHOUT "STRAIN-HARDENING"

The success of the hybrid model with $n=1$ has implications for the theoretical framework of the grain-to-grain shearing mechanism that Buckingham develops in Ref. 4. The material exponent is defined as $n=E_p/\theta_p$, where E_p is the elastic response of the micro-asperity contact and θ_p reflects the nonlinear response of the viscous damping at the contact. Thus $n=1$ means that elastic and viscous responses

TABLE II. Material exponents and compressional rigidity coefficients used in the hybrid model for each sediment. The parameters were determined from a fit to the data allowing both parameters to be free in the first case and fixing $n=1$ in the second case.

Sediment	Fit with two free parameters		Fit with one free parameter	
	n	γ_p (Pa)	n	γ_p (Pa)
SAX99	1.088	41.96	1.00	146.82
Water/glass beads	0.915	182.94	1.00	54.79
Si100/glass beads	1.059	57.89	1.00	115.55

are equal, which is extremely unlikely since they are unrelated to one another. This may also mean, however, that the "strain-hardening" is actually absent from the grain-to-grain interactions, i.e., $\theta_p=0$, and the viscosity of the fluid film remains linear during the micro-sliding event.

To understand why this may be the case, consider the differential equation of motion for the micro-sliding event given in Eq. (16) in Ref. 4,

$$\frac{1}{E}\frac{d\chi}{dt} + \frac{1}{\xi(t)}\chi = \varepsilon\,\delta(t). \tag{26}$$

If the dashpot, $\xi(t)$, is no longer a function of time, the strain-hardening coefficient θ_p is zero and $\xi(t)=\xi_0$. The solution to the equation of motion becomes

$$\chi = \chi_0 e^{-(E/\xi_0)t}\,\mathrm{sgn}[\varepsilon\,\delta(t)]. \tag{27}$$

The time-dependent part of the solution is the pulse shape function, or MIRF,

$$h(t) = \frac{E}{\xi_0}e^{-(E/\xi_0)t}, \tag{28}$$

and the mean pulse amplitude becomes

$$\bar{a} = \chi_0\frac{\xi_0}{E}. \tag{29}$$

Using these new expressions, the stress associated with radial and translational sliding given in Eqs. (4) and (5) is unchanged, but the constants now become

$$\lambda_p = \chi_{0p} \nu_p \bar{b} t_p, \qquad \eta_s = \chi_{0s} \nu_s \bar{b} t_s, \tag{30}$$

where $\bar{b}$ is a proportionality constant defined in Ref. 4, the stress relaxation time constants are

$$t_p = \frac{\xi_{0p}}{E_p}, \qquad t_s = \frac{\xi_{0s}}{E_s} \tag{31}$$

and the MIRFs are now

$$h_p(t) = t_p^{-1} e^{-t/t_p} \tag{32}$$

and

$$h_s(t) = t_s^{-1} e^{-t/t_s}. \tag{33}$$

We can now obtain expressions for the shear and bulk moduli by taking the Fourier transforms of the MIRFs,

$$H_p(\omega) = \frac{1}{1 - i\omega t_p}, \qquad H_s(\omega) = \frac{1}{1 - i\omega t_s} \tag{34}$$

and substituting these into the expressions for the shear and bulk moduli in Eqs. (9) and (12). The new frequency dependent moduli then become

$$\mu_b(\omega) = \frac{\eta_s}{1 + i\dfrac{1}{\omega t_s}} \tag{35}$$

and

$$K_b(\omega) = \frac{\lambda_p}{1 + i\dfrac{1}{\omega t_p}}. \tag{36}$$

It is reasonable to assume that the spring constant, E, associated with the compressibility of the grain asperities and hence related to the elastic properties of the grain ($>10^9$ Pa), is much larger than the viscosity of the fluid flim, ξ_0, which should be related to the viscosity of the bulk fluid (<0.1 Pa s). If this is the case then $\omega t_p \ll 1$ and the bulk modulus is approximately

$$K_b(\omega) = -i\gamma_p \omega, \tag{37}$$

which has the same form as Eq. (16) when $n=1$ and with $\gamma_p = \lambda_p t_p$. The data model comparisons given in Sec. IV are therefore applicable for this linear form of the grain-to-grain shearing model.

In the Appendix, we show that for $K_b, \mu_b \ll K_r, K_f$, the hybrid model can be expressed as

$$k^2 = \frac{\omega^2 \rho_{\text{eff}}}{K(\omega)} \tag{38}$$

where

$$K(\omega) = H_{\text{eff}} + K_b(\omega) \frac{\left(1 - \dfrac{2H_{\text{eff}}}{K_r}\right)(\tilde{\rho} - \rho_f) + \rho_f - \rho_{\text{eff}}}{\rho + \tilde{\rho} - 2\rho_f}$$

$$+ \mu_b(\omega) \frac{\tilde{\rho} - \rho_{\text{eff}}}{\rho + \tilde{\rho} - 2\rho_f}. \tag{39}$$

This expression extends the EDFM to include the effects of a weak frame stiffness or loss in the effective bulk modulus of the medium. As expected, the frequency dependent modulus $K(\omega)$ reduces to H_{eff} and the wave number reduces to the EDFM result when the frame moduli are set to zero. Note that this expression does not depend on the form of the bulk moduli only that they are much less than the pore fluid and grain material moduli. This expression is a good approximation to the full hybrid model solution over most of the frequency range and begins to deviate at high frequencies. For the SAX99 parameters and the rigidity coefficients in Table II, the error in the attenuation at 100 kHz is 1.2% which increases to 7% at 1 MHz. The sound speed error is 0.0002% at 100 kHz which increases to 0.02% at 1 MHz.

The form of the frequency dependent bulk modulus in Eq. (36) is identical to the Biot model with grain contact squirt flow and shear drag (BICSQS model) proposed by Chotiros and Isakson[14] with the exception of a purely real asymptotic frame bulk modulus. For the shear modulus, the expressions differ between the two models, but the asymptotic values are the same. In their fits to the SAX99 data, the frequency dependent bulk modulus is responsible for fitting the lowest frequency SAX99 data, which we have not attempted to fit. It is the shear modulus which largely determines the fit to the high frequency attenuation data. This ignores the shear speed and attenuation data at 1 kHz collected during SAX99 which, similar to Buckingham's model and the hybrid model, determines the two free parameters of the frequency dependent shear modulus. Applying the BIC-SQS model to the shear data using the parameters given in Table I, yields a relaxation frequency of $f_\mu = 2$ Hz significantly lower than the 56 kHz value found by the Chotiros and Isakson. Note that this fitting procedure does not treat the permeability as a free parameter as in Ref. 14, but we do not feel that the permeability allows sufficient freedom to fit both the shear data point and the high frequency compressional attenuation. However, if the requirement to fit the lowest frequency sound speed data points is relaxed, the parameters in the bulk modulus can be adjusted to fit the compressional attenuation data as in the linear hybrid model above. This leaves open the question of which, if either, of these two models is physically correct. Both will yield valid fits that are indistinguishable and a definitive test will require comparisons of the model parameters to measurable properties of the sediment. However neither model has been developed to the point where this comparison can be performed.

VI. CONCLUSIONS

The purpose of this experiment was to make sound speed and attenuation measurements in a simple glass-bead sediment and to observe how these measurements changed as the viscosity of the pore fluid was increased by two orders of magnitude. Several conclusions can be drawn from the measurements and the data/model comparisons with Biot-Stoll theory:

1. Biot theory correctly predicted that the frequency range

over which the most significant dispersion occurs should shift into the frequency range of our measurement system as the viscosity increased.

2. While the frequency dependence of the attenuation differed from the predictions of Biot theory, the theory was also able to predict how the attenuation scales with viscosity.

3. From this we conclude that the relative motion of the pore fluid through the sediment frame plays a significant role in both the dispersion and the attenuation, in contrast to the conclusion reached by Buckingham, and must be taken into account to properly model sound propagation through sand sediments.

4. Measurements of attenuation in the SAX99 sediment indicated that the high frequency attenuation follows a linear frequency dependence as opposed to the $f^{1/2}$ dependence which is predicted by Biot theory. It was suggested that this excess attenuation might have been due to scattering from shells or shell fragments. This linear frequency dependence was also observed in the glass-bead sediments in the absence of scatterers suggesting that the mechanism for this excess attenuation is more fundamental to a granular medium such as losses at the grain contacts.

Since Buckingham's grain-to-grain shearing model could be *adjusted* to fit the frequency dependence of the attenuation, while Biot theory was able to *predict* the sound speed dispersion in all three sediments, we constructed a hybrid model which incorporated grain-to-grain shearing physics into the frame moduli in Biot theory. The resulting theory captured both the motion of the viscous pore fluid and the interactions of the grain contacts. The process of combining these models did not explicitly violate the physics behind either of the models and in fact the incorporation of complex, frequency dependent moduli into Biot theory was first suggested by Biot himself.

Fits of the hybrid model to the data implied that the exponent in Buckingham's theory should be $n=1$. This further implies that the viscous response of the asperity contacts can be treated as linear which means that "strain-hardening" suggested by Buckingham is absent. With this assumption, a linear version of grain-to-grain shearing can be incorporated into Biot theory to give the proper frequency dependence for the complex moduli. The resulting bulk modulus has a form that is similar to that proposed in the BICSQS model, however at this point in the development of both models, neither has been developed to the point where the material properties of the sediment can be compared to the relevant parameters in the models. Therefore, the frame loss embodied in Eqs. (15) and (16) may be due to neither, one, or both of these mechanisms. Regardless, we believe the experiment/model comparisons given here imply that Biot's suggestion of adding frame loss phenomenologically using the form of Eqs. (15) and (16) is viable.

Finally, Eq. (38) can be used with Eqs. (15), (16), and (39) to model unconsolidated sand sediments as a fluid, while at the same time accounting for frame loss, with only a small loss in accuracy over the full model developed in Secs.

IV and V. We believe this simplified fluid model can be useful for application oriented work as well as basic research where porous media effects are being studied within a larger context, e.g., long range, shallow water propagation and scattering.

APPENDIX: EXTENDING THE EDFM TO INCLUDE FRAME MODULI

Here, we revisit Biot's equations in order to derive an approximate solution similar to the effective density fluid model (EDFM) which incorporates the frame bulk and shear moduli to second order. Using the potentials[25] $(\Phi_s, \Phi_f, \Psi_s, \Psi_f)$ defined in terms of the displacement vectors of the skeletal frame $\mathbf{u}$ and the fluid $\mathbf{U}$ as

$$
\begin{aligned}
\mathbf{u} &= \nabla\Phi_s + \nabla \times \mathbf{\Psi}_s, \\
\beta(\mathbf{u} - \mathbf{U}) &= \nabla\Phi_s + \nabla \times \mathbf{\Psi}_s.
\end{aligned}
\tag{A1}
$$

Biot's equations for the scalar potentials are given as (using a plane wave solution $\exp[i(\mathbf{k}\cdot\mathbf{r} - \omega t)]$)

$$
-k^2 H\Phi_s + k^2 C\Phi_f = -\rho\omega^2\Phi_s + \rho_f\omega^2\Phi_f,
\tag{A2}
$$

$$
-k^2 C\Phi_s + k^2 M\Phi_f = -\rho_f\omega^2\Phi_s + \frac{\omega^2\alpha\rho_f}{\beta}\Phi_f + \frac{i\omega F\eta}{\kappa}\Phi_f.
\tag{A3}
$$

In Eqs. (A2) and (A3), k is the wave number, α is the structure constant or tortuosity, η is the pore fluid viscosity, ρ_s is the sediment particle density, ρ_f is the pore fluid density, and

$$
k = \frac{\omega}{c}, \quad \omega = 2\pi f,
$$

$$
\rho = \beta\rho_f + (1-\beta)\rho_s,
$$

$$
H = \frac{(K_r - K_b)^2}{D - K_b} + K_b + \frac{4\mu}{3},
$$

$$
C = \frac{K_r(K_r - K_b)}{D - K_b},
\tag{A4}
$$

$$
M = \frac{K_r^2}{D - K_b},
$$

$$
D = K_r\left[1 + \beta\left(\frac{K_r}{K_f} - 1\right)\right] = \frac{K_r^2}{H_{\text{eff}}},
$$

$$
H_{\text{eff}} = \left(\frac{(1-\beta)}{K_r} + \frac{\beta}{K_f}\right)^{-1},
\tag{A5}
$$

where f is the frequency, c is the sound speed, ρ is the total mass density, K_r is the bulk modulus of the individual sediment grains, K_f is the bulk modulus of the pore fluid, and H_{eff} corresponds to the bulk modulus of the sediment found

for the EDFM when $K_b = \mu_b = 0$ in H, C, and M. The parameter F represents the deviation from Poiseuille flow as the frequency increases and is discussed in detail in Ref. 12 and the references cited therein.

If $K_b, \mu_b \ll K_r, K_f$, Biot's coefficients can be approximated to second order in the frame moduli as

$$H \approx H_{\text{eff}} + \left(1 - \frac{2H_{\text{eff}}}{K_r}\right)K_b + \frac{4\mu_b}{3},$$

$$C \approx H_{\text{eff}}\left(1 - \frac{K_b}{K_r}\right),$$

$$M \approx H_{\text{eff}}. \tag{A6}$$

Note that this approximate form for the coefficients produces the same results as the full coefficients for the "jacketed" and "unjacketed" tests introduced by Biot[9,26] to help understand the significance of these coefficients. Using Eqs. (A2) and (A3) with the approximate coefficients given in Eq. (6), the wave number for longitudinal waves can be found from

$$k^2 = \frac{-b\left(1 \pm \sqrt{1 - \frac{4ac}{b^2}}\right)}{2a}, \tag{A7}$$

where the $(-)$ sign corresponds to the fast wave, the $(+)$ sign corresponds to the slow wave, and

$$a \approx H_{\text{eff}}(K_b + 4\mu_b/3),$$

$$b \approx H_{\text{eff}}(2\rho_f - \rho - \tilde{\rho})$$
$$- \left[\left(1 - \frac{2H_{\text{eff}}}{K_r}\right)\tilde{\rho} + \frac{2H_{\text{eff}}}{K_r}\rho_f\right]K_b - \frac{4\mu_b}{3},$$

$$c = \tilde{\rho}\rho - \rho_f^2,$$

$$\tilde{\rho} = \frac{\alpha\rho_f}{\beta} + \frac{iF\eta}{\kappa\omega}. \tag{A8}$$

Since the frame moduli are assumed to be much smaller than the material moduli, the square root can the approximated by its Taylor series, where terms higher than second order are neglected. The resulting approximate wave number for the fast wave can be written as

$$k^2 = \frac{\omega^2 \rho_{\text{eff}}}{K}, \tag{A9}$$

where ρ_{eff} is the complex effective density for the EDFM and defined in Ref. 12. The effects of the frame moduli are contained in the effective bulk modulus

$$K = H_{\text{eff}} + K_b \frac{\left(1 - \frac{2H_{\text{eff}}}{K_r}\right)(\tilde{\rho} - \rho_f) + \rho_f - \rho_{\text{eff}}}{\rho + \tilde{\rho} - 2\rho_f}$$
$$+ \mu_b \frac{\tilde{\rho} - \rho_{\text{eff}}}{\rho + \tilde{\rho} - 2\rho_f}. \tag{A10}$$

For the complex moduli used in Ref. 12, the error in the sound speed relative to the full Biot model using this approximation is less than 0.02% and the error in the attenua-tion is less than 1% over a frequency range extending from 100 Hz to 10 MHz.

[1] E. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

[2] M. Buckingham, "Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments," J. Acoust. Soc. Am. **102**, 2579–2596 (1997).

[3] M. Buckingham, "Theory of compressional and shear waves in fluid-like marine sediments," J. Acoust. Soc. Am. **103**, 288–299 (1998).

[4] M. Buckingham, "Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am. **108**, 2796–2815 (2000).

[5] A. Kibblewhite, "Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data," J. Acoust. Soc. Am. **86**, 716–738 (1989).

[6] M. Biot, "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**, 1254–1264 (1962).

[7] M. Biot, "Mechanics of deformation and acoustic propagation in porous media," J. Appl. Phys. **33**, 1482–1492 (1962).

[8] R. Stoll, "Marine sediment acoustics," J. Acoust. Soc. Am. **77**, 1789–1799 (1985).

[9] R. Stoll, *Sediment Acoustics, Lecture Notes on Earth Sciences* (Springer, Berlin 1989).

[10] E. Thorsos, K. Williams, N. Chotiros, J. Christoff, K. Commander, C. Greenlaw, D. Holliday, D. Jackson, J. Lopes, D. McGehee, J. Piper, M. Richardson, and D. Tang, "An overview of SAX99: Acoustic measurements," IEEE J. Ocean. Eng. **26**, 4–25 (2001).

[11] M. Richardson, K. Briggs, L. Bibee, P. Jumars, W. Sawyers, D. Albert, R. Bennett, T. Berger, M. Buckingham, N. Chotiros, P. Dahl, N. Dewitt, P. Fleischer, R. Flood, C. Greenlaw, D. Holliday, M. Hulbert, M. Hutnak, P. Jackson, J. Jaffe, H. Johnson, D. Lavoie, A. Lyons, C. Martens, D. McGehee, K. Moore, T. Orsi, J. Piper, R. Ray, A. Reed, R. Self, J. Schmidt, S. Schock, F. Simonet, R. Stoll, D. Tang, D. Thistle, E. Thorsos, D. Walter, and R. Wheatcroft, "An overview of SAX99: Environmental considerations," IEEE J. Ocean. Eng. **26**, 26–54 (2001).

[12] K. L. Williams, D. Jackson, E. Thorsos, D. Tang, and S. Schock, "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media," IEEE J. Ocean. Eng. **27**, 413–428 (2002).

[13] P. Seifert, B. Kaelin, and L. Johnson, "Effect on ultrasonic signals of viscous pore fluids in unconsolidated sand," J. Acoust. Soc. Am. **106**, 3089–3094 (1999).

[14] N. Chotiros and M. Isakson, "A broadband model of sandy ocean sediments: Biot-Stoll with contact squirt flow," J. Acoust. Soc. Am. **116**, 2011–2022 (2004).

[15] F. Bowden and D. Tabor, *Friction and Lubrication of Solids* (Oxford University Press, New York, 2000).

[16] J. Bear, *Dynamics of Fluids in Porous Media* (Dover, New York, 1988).

[17] J. Berryman, "Effective conductivity by fluid analogy for a porous insulator filled with a conductor," Phys. Rev. B **27**, 7789–7792 (1983).

[18] J. Berryman, "Confirmation of Biot's theory," Appl. Phys. Lett. **34**, 382–384 (1980).

[19] M. Buckingham and M. Richardson, "On tone-burst measurements of sound speed and attenuation in sandy marine sediments," IEEE J. Ocean. Eng. **27**, 429–454 (2002).

[20] K. Williams, "An effective density fluid model for acoustic propagation in sediments derived from Biot theory," J. Acoust. Soc. Am. **110**, 2276–2281 (2001).

[21] R. D. Costley and A. Bedford, "An experimental study of acoustic waves in saturated glass beads," J. Acoust. Soc. Am. **83**, 2165–2174 (1988).

[22] E. Kjartansson, "Constant Q-wave propagation and attenuation," J. Geophys. Res. **84**, 4737–4748 (1979).

[23] A. Turgut, "An investigation of causality for Biot models by using Kramers-Krönig relations," in *Shear Waves in Marine Sediments*, edited by J. Hovem, M. Richardson, and R. Stoll (Kluwer, Dordrecht, 1991) pp. 295–304.

[24] B. Brunson and R. Johnson, "Laboratory measurements of shear wave attenuation in saturated sand," J. Acoust. Soc. Am. **68**, 1371–1375 (1980).

[25] R. Stoll and T. Kan, "Reflection of acoustic waves at a water-sediment interface," J. Acoust. Soc. Am. **70**, 149–156 (1981).

[26] M. Biot and D. Willis, "The elastic coefficients of the theory of consolidation," J. Appl. Phys. **24**, 594–601 (1962).

An investigation into the effects of underwater piling noise on salmonids

Jeremy R. Nedwell
Subacoustech Ltd., Chase Mill, Winchester Road, Bishop's Waltham, Hampshire SO32 1AH, United Kingdom

Andrew W. H. Turnpenny
Jacobs Babtie Aquatic, Fawley, Southampton, Hampshire SO45 1TW, United Kingdom

Jonathan M. Lovell
University of Plymouth, Drake Circus, Plymouth PL4 8AA, United Kingdom

Bryan Edwards
Subacoustech Ltd., Chase Mill, Winchester Road, Bishop's Waltham, Hampshire SO32 1AH, United Kingdom

Underwater piling was undertaken in 2003 in Southampton Water on the South Coast of England. Monitoring was simultaneously undertaken of the waterborne sound from impact and vibropiling and its effects on brown trout in cages at increasing distances from the piling. Brown trout (*Salmo trutta*) were used as a model for salmon (*Salmo salar*), which were the species of interest but were not readily available. No obvious signs of trauma that could be attributed to sound exposure were found in any fish examined, from any of the cages. No increase in activity or startle response was seen to vibropiling. Analysis using the dB_{ht} metric indicated that the noise at the nearest cages during impact piling reached levels at which salmon were expected to react strongly. However, the brown trout showed little reaction. An audiogram of the brown trout was measured by the Auditory Brainstem Response method, which indicated that the hearing of the brown trout was less sensitive than that of the salmon. Further analysis indicated that this accounted for the relative lack of reaction, and demonstrated the importance of using the correct species of fish as a model when assessing the effect of noise. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2335573]

I. INTRODUCTION

It has been documented that the driving of piles in water causes high levels of underwater noise. Measurements made of piling using piles of 0.7 m diameter on the River Arun in southern England indicated a peak-to-peak Source Level of 191 dB $re1$ μPa @ 1 m (Nedwell *et al.*, 2002). Measurements made of the noise from piling using piles of 4.3 m diameter during the construction of offshore windfarms in the UK indicated a peak-to-peak Source Level of 260 dB $re1$ μPa @ 1 m for 5 m depth, and 262 dB $re1$ μPa @ 1 m at 10 m depth, associated with a Transmission Loss given by $22 \log(R)$, where R is the range (Nedwell *et al.*, 2003b). Such levels of sound may have an effect on marine species. A report for Caltrans on piling in San Francisco Bay indicated mortality to several species of fish at ranges of up to 50 m (Abbott, 2005). Intense sound from piling may also have an effect on hearing. McCauley *et al.* (2003) reported damage to the ears of a pink snapper caused by the impulsive noise of an airgun having a peak-to-peak Source Level of 222.6 dB $re1$ μPa @ 1 m and deployed at a range of 400–800 m from the caged fish. Enger (1981) reported damage to the ciliary bundles on the sensory cells of the inner ear of cod caused by sound at frequencies of 50 to 400 Hz at a level of 180 dB.

In September 2003, piling was undertaken in Southampton Water, a shallow inlet on the South Coast of England, adjacent to the quayside of a ferry terminal. There was concern that the construction work might adversely impact local fish populations, and, in particular, the migration of salmon up Southampton Water into the River Test, a Special Area of Conservation (SAC). A limit of 90 dB_{ht}(*Salmo salar*) at half-channel width was set as a noise limit by the regulator (the UK Environment Agency), with the intent of leaving half of the channel open for migration; the dB_{ht} metric is discussed in Sec. III B. Consequently, monitoring was undertaken of the waterborne sound generated by the piling operations. Simultaneously, the opportunity was taken to monitor the effects of the noise on fish in cages at increasing distances from the piling.

II. FIELD WORK

A. Location of measurements

The piling was carried out adjacent to a quay, in otherwise open water conditions. A total of ten piles, four 914 mm in diameter and six 508 mm in diameter, were driven. Vibrodriving was used for driving the piles, the duration being approximately 20 min per pile. At the end of the operation 3 of the 10 piles were driven to final depth by impact driving

for dynamic test purposes, the duration of each test being between one and three minutes, with approximately 20 blows during the period.

B. Fish response monitoring

Hydrophones and fish cages for reaction monitoring were located at 30, 54, 96, 234, and 417 m from the piling. The cages were 1 m cubes, made from a mild steel angle and covered with plastic mesh of 25 mm square aperture and suspended 2.5 m below the water surface. A closed circuit television camera was placed in each cage. A control cage was suspended in a dock 6 km from the site of the piling, where the piling noise had fallen below background levels.

Although the species of interest was salmon (*Salmo salar*), suitable fish were not available and instead farmed brown trout (*Salmo trutta*) were substituted as a close relative, which it was thought would react similarly. As is noted in Sec. III B, this assumption proved to be invalid.

C. Fish behavioral reactions

The video recordings were reviewed to identify any changes in behavior that might have resulted from the piling noise. These data were reviewed "blind," with the reviewer unaware of the condition of exposure. Two types of behavior were investigated.

(i) Startle reactions, which are here defined as sudden C-shaped flexure of the fish's body, described by Blaxter and Hoss (1981). The analysis of the startle reactions was based on a frame-by-frame inspection of the video at the start-up instant of each vibropiling session and for the next 5 s;

(ii) Fish activity level; captive fish that are exposed to irritating stimuli commonly show "milling behavior," in which the fish swim faster and make random turns. These were measured by counting the number of times fish entered the camera's field of view within a two-minute observation period.

Due to the fish being caged avoidance reaction (i.e., fleeing the noise) could not be directly observed, and the activity level was used as a measure of the behavioral effect. It may be commented, however, that human experience would indicate that the avoidance of noise may well occur at lower levels of noise than an increase in activity, and startle probably occurs at relatively high levels of noise. Since the caged fish could not flee, the experiment was imperfect in that it probably did not demonstrate the lowest level of sound at which avoidance of the piling noise occurred.

1. Reactions during vibropiling

No startle response was seen for any of the piles driven by vibropiling.

Table I gives the activity level observations for the 2 min period prior to piling and then for the first 2 min during vibropiling; they are seen to have remained similar before and after the start of piling. The control and event activity levels were compared using the nonparametric Mann-Whitney U-test (Campbell, 1974) with the null hypothesis that activity levels were not significantly different at the

TABLE I. Fish activity statistics (No. of movements per 2 min period) for vibropiling.

| Range (m) | Mann-Whitney U test | | | | |
	Before	During	U	Z	P
30	147.5	152.5	69.5	0.1443	0.8852
54	150.5	149.5	71.5	−0.028	0.9769

$P = 0.05$ level; it was found that there was no significant difference in activity level following the commencement of vibropiling ($P = 0.001$).

2. Reactions during impact piling

One pile of 914 mm diameter and two of 508 mm were impact driven. No startle reactions were observed at any of the five locations.

Activity levels recorded before and during the impact piling sessions are given in Table II. Fish activity is seen to have remained similar in most cases before and after the start of impact piling. The Mann-Whitney U test shows that there was no significant difference in activity level following the commencement of pile driving ($P = 0.001$) in cages at 30, 96, 234, and 417 m; a 36% increase in fish activity level was, however, observed in the cage at 54 m, which was significant at the $P = 0.05$ level.

III. LABORATORY STUDIES

A. Electron microscope examination of the inner ear of the trout (*Salmo trutta*)

The fish were initially examined under an optical microscope after the end of the piling operation for any evidence of swim bladder rupture, eye haemorrhage, or eye embolism as a result of exposure to the sound from the piling; none was found. The saccule, the primary auditory region of the fish ear (Popper and Fay, 1993), was investigated for evidence of trauma to the inner ear ultrastructure. Five fish from each of the cages were examined for ultrastructural damage to the hearing organs; the examiner was unaware of which fish came from which cage. A typical Scanning Electron Micrograph (SEM) of saccular hair cells from *S. trutta* is presented in Fig. 1, which is from a central region of the macula. The image shows apical hair cell bundles with an anterior positioned kinocilia, surrounded by approximately 40 cilia arranged in 4 to 5 consecutively shorter rows in a format common to many fish species (e.g., Platt and Popper, 1981; Lovell *et al.*, 2005). Each hair cell is separated from

TABLE II. Fish activity statistics (No. of movements per 2 min period) for impact piling.

| Distance | Mann-Whitney U test | | | | |
	Before	After	U	Z	P
30 m	9	12	3	−0.654	0.513
54 m	6	15	0	−1.96	0.0495
96 m	12	9	3	0.654	0.512
234 m	6.5	3.5	0.5	1.16	0.245
417 m	12	9	3	0.654	0.512

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Nedwell *et al.*: Effects on fish of piling noise   2551
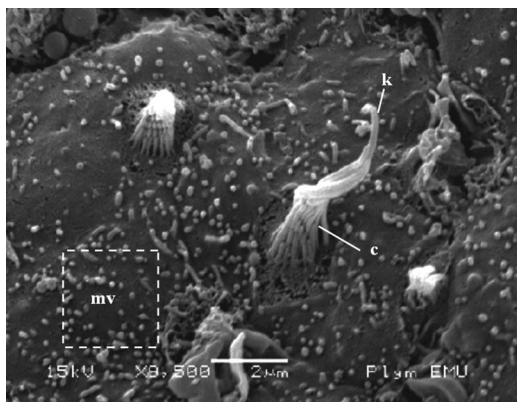
FIG. 1. Scanning Electron Micrograph (SEM) of saccular hair cells from *S. trutta*, from a central region of the macula. Fish from cage number 1. k. kinocilia, c. cilia, mv. microvilli.

neighboring cells by an area of epithelium bearing short (<1 μm) microvilli. No obvious signs of trauma that could be attributed to sound exposure were found in any fish examined, from any of the cages.

B. Fish audiograms

1. Quantifying the effects of noise on species

The response of marine animals to a given sound is dependent on the species since these vary greatly in hearing sensitivity and frequencies range. A metric has been described (Nedwell *et al.*, 2003a), (Nedwell *et al.*, 2004) termed the $dB_{ht}(Species)$ scale, which may be regarded as a generalization of the dB(A) metric used for the rating of the behavioral effects of noise on humans and hence that is indicative of how much a species will be affected by that sound. It is an estimate of the level that the sound is above the hearing threshold of the species and hence is an indication of the likely perception by, or "loudness" of the sound to, that species. Initial research with fish (Nedwell, in preparation) indicates that at levels of 90 dB or more above the species' threshold of hearing [i.e., of 90 $dB_{ht}(Species)$ or more] strong avoidance of sound occurs.

2. Inadequacies of brown trout as a model for salmon

At the time of the experimental measurements, it was noted that, despite the piling reaching levels of noise at which salmon, *S. salar,* were expected to react [i.e., in excess of 90 $dB_{ht}(S. salar)$], the brown trout, *S. trutta,* were showing no obvious signs of reaction. It was suspected that this was a result of the hearing of the latter being less sensitive. Since there was no audiogram available for the brown trout, this was determined by the Auditory Brainstem Response (ABR) method, and compared with previous published audiograms of salmon (Hawkins and Johnstone, 1978), in order to attempt to explain this unexpected observation. It should be noted, however, that the latter audiogram was obtained by the behavioral method, which in humans has been noted to yield slightly different results from ABR audiograms.
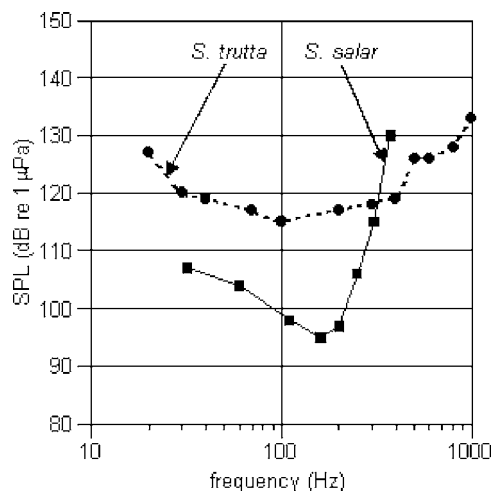
FIG. 2. Audiograms for the brown trout (*Salmo trutta*) and the salmon (*Salmo salar*). Salmon audiogram from Hawkins and Johnstone (1978).

3. The ABR method

The ABR method was similar to that described by Kenyon *et al.* (1998). Fish were held in a cradle in a water tank and a pair of electrodes placed cutaneously on the cranium such that they spanned the VIIIth nerve, recording the evoked auditory response to sound. The sound stimulus of several periods of a sine wave at a given frequency was generated by two underwater projectors within a water tank. Responses were averaged over 2000 presentations. The level of the stimulus was reduced until the stimulus trace was no longer discernible in the response trace. The sound level at which the stimulus trace was just discernible was taken as the threshold level for that frequency. The procedure was carried out using four fish at all of the frequencies tested, and 24 fish at frequencies between 300 and 1000 Hz.

Figure 2 shows the resulting *S. trutta* audiogram, as well as that for *S. salar* from Hawkins and Johnstone (1978). First, it should be noted that the latter audiogram was obtained by the behavioral method, which may give slightly different results from the ABR method. Nevertheless, it may be seen that, despite being a close relative with morphologically similar hearing, the audiogram of the brown trout is significantly different from that of the salmon. It is less sensitive, and broader and flatter in frequency response, with effective hearing from 30 Hz to above 1 kHz. It was concluded that the common assumption that closely related species will have similar hearing abilities is not reliable. Consequently, the data recorded during the piling were reprocessed using the brown trout audiogram to yield $dB_{ht}(S. trutta)$ values in addition to the $dB_{ht}(S. salar)$ values recorded on-site at the time of the noise measurements; these data are presented in Sec. IV.

IV. ANALYSIS OF NOISE MEASUREMENTS

A. Introduction

The piling was recorded using Brüel & Kjær hydrophones calibrated to International Standards, and conditioned by Brüel & Kjær charge amplifiers, before being digitized,
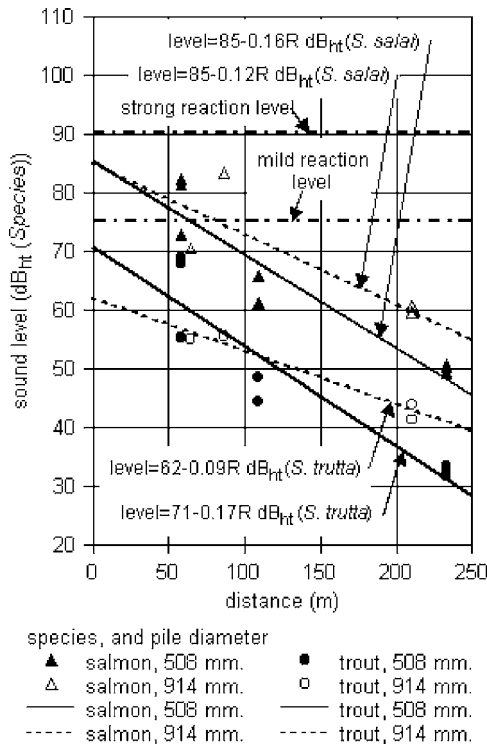
FIG. 3. The average peak-to-peak sound pressure level as a function of range, for the impact piling case, in dB$_{ht}$(*Species*) units, for two diameters of pile.

archived, and analyzed using a program written in the National Instruments LabVIEW environment that calculated one second dB$_{ht}$(*Species*) values.

B. Results

It was found that the unweighted Source Level of the impact piling of the 508 mm diameter pile was 193 dB *re*1 μPa @ 1 m, with a linear Transmission Loss rate of 0.13 dB per meter, and for the 914 mm diameter pile the Source Level was 201 dB *re*1 μPa @ 1 m and the Transmission Loss 0.13 dB per meter.

Figure 3 presents the dB$_{ht}$ levels of the impact piling as a function of range. The levels recorded at each of the measurement positions are indicated on the figure; for each of the two pile diameters driven the values of the noise in dB$_{ht}$(*Species*) units for both species have been plotted. It will be noted that the levels are different for the same data for the two different species since they have different hearing ability and frequency range of hearing. The figure may therefore be interpreted as indicating "loudness" for the two species as a function of range. Also illustrated in the figure is the 90 dB$_{ht}$(*Species*) limit at which it is believed noise becomes intolerable.

For each set of results a least squares best fit of the data has been calculated, the results being

$$SPL = 85 - 0.12R \ dB_{ht}(S. \ salar) \quad \text{and} \quad SPL = 62$$

$$- 0.09R \ dB_{ht}(S. \ trutta), \tag{1}$$

for the 914 mm diameter pile, and

$$SPL = 85.4 - 0.16R \ dB_{ht}(S. \ salar) \quad \text{and} \quad SPL = 70.7$$

$$- 0.17R \ dB_{ht}(S. \ trutta), \tag{2}$$

for the 508 mm diameter pile.

In general, the dB$_{ht}$(*S. salar*) levels are higher than the dB$_{ht}$(*S. trutta*) levels. This is to be expected, as the hearing of the salmon is more sensitive than that of the brown trout. It is noticeable that there is a significant difference in Transmission Loss for the case of *S. trutta* for the two pile diameter cases. It is thought likely that this might have been due to the fit of Source Level and Transmission Loss to the data being poor in the case of the 914 mm diameter pile.

In respect of the results for *S. trutta,* it may be seen that both the measured and the estimated levels near to the piling are well below the level at which a mild reaction would be expected to occur. In other words, no reaction would be expected by *S. trutta* at any range from the piling. By comparison, the results for *S. salar* would indicate that a mild reaction would be expected at a range of about 60–80 m, and an increasingly stronger reaction as this range was reduced. These results may explain the relative lack of a reaction to the piling from the brown trout *S. trutta*, and indicate the importance of using the correct species of fish as a model when assessing the effect of noise.

V. CONCLUSION

In summary we have the following.

(1) Observations were made of caged brown trout (*S. trutta*) during vibropiling of four 914 mm and six 508 mm diameter piles. The brown trout was used as a model for salmon (*S. salar*), which were not readily available.

(2) No reaction to vibropiling was noted at any of the cages. No startle reactions were observed for the impact driving. However, there was a 36% increase in the fish activity level observed in the cage at 54 m, which was significant at the $P = 0.05$ level, although in the two cages nearest to the piling (at 30 and 54 m), no significant difference in activity level was detected.

(3) Despite the noise from the piling at the nearest cages reaching levels at which salmon, *S. salar,* were expected to react strongly, the brown trout, *S. trutta,* showed little reaction. An audiogram of the brown trout measured by the Auditory Brainstem Response method indicated that the hearing of the brown trout was less sensitive than that of the salmon. Analysis using the dB$_{ht}$ metric indicated that this could account for the relative lack of reaction, and indicated the importance of using the correct species of fish as a model when assessing the effect of noise.

Abbott, R. (**2005**). "Fisheries and hydroacoustic monitoring program compliance report addendum." San Francisco – Oakland Bay Bridge East Span Seismic Safety Project. May 2005, Office of Biological Sciences and Permits, Caltrans District 4.

Blaxter, J. H. S., and Hoss, D. E. (1981). "Startle responses in herring: the effect of sound stimulus frequency, size of fish and selective interference with the acousto-lateralis system," J. Mar. Biol. Assoc. **61**, 871–879.

Campbell, R. C. (**1974**). *Statistics for Biologists* (Cambridge University Press, Cambridge).

Enger, P. S. (**1981**). "Frequency discrimination in teleosts—central or peripheral?," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay. (Springer, New York), pp.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Nedwell *et al.*: Effects on fish of piling noise   2553

243–255.

Hawkins, A. D., and Johnstone, A. D. F. (**1978**). "The hearing of the Atlantic salmon, *Salmo salar*," J. Fish Biol. **13**, 655–673.

Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). "A comparative study of hearing ability in fishes: the auditory brainstem response approach," J. Comp. Physiol., A **182**, 307–318.

Lovell, J. M., Findlay, M. M., Moate, R. M., and Pilgrim, D. A. (**2005**). "The polarization of inner ear ciliary bundles from a scorpaeniform fish," J. Fish Biol. **66**, 836–846.

McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). "High intensity anthropogenic sound damages fish ears," J. Acoust. Soc. Am. **113**, 638–642.

Nedwell, J. R. (in preparation).

Nedwell, J. R., Turnpenny, A. W. H., and Edwards, B. (**2002**). "Piling on the River Arun; implications for salmon migration," *American Fisheries Society 132nd Annual Meeting*, 18–22 August 2002, Hyatt Regency Hotel, Baltimore, MD.

Nedwell, J. R., Turnpenny, A. W. H., and Lambert, D. R. (**2003a**). "Objective design of acoustic fish deflection systems," *Proceedings of the US EPA Symposium on Cooling Water Intake Technologies to Protect Aquatic Organisms*, 6–7 May, 2003, Hilton Crystal City at National Airport, Arlington, VA.

Nedwell, J. R., Langworthy, J., and Howell, D. (**2003b**). "Assessment of sub-sea acoustic noise and vibration from offshore wind turbines and its impact on marine wildlife; initial measurements of underwater noise during construction of offshore windfarms, and comparison with background noise," Subacoustech Report No. 544 R 0424 (for the Crown Estate), May 2003.

Nedwell, J. R., Langworthy, J., and Howell, D. (**2004**). "The dB$_{ht}$; a methodology for evaluating the behavioral effects of underwater noise and some examples of its use," *Proceedings of the Symposium on Bio-Sonar Systems and Bio-Acoustics*, Institute of Acoustics, 16 September 2004, Loughborough University, UK.

Platt, C., and Popper, A. N. (**1981**). "Fine structure and function of the ear," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. A. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 3–36.

Popper, A. N., and Fay, R. R. (**1993**). "Sound detection and processing by fish: critical review and major research questions," Brain Behav. Evol. **41**, 14–38.

Mean grain size mapping with single-beam echo sounders

Paul A. van Walree[a)] and Michael A. Ainslie

TNO Defence, Security and Safety, Oude Waalsdorperweg 63, P.O. Box 96864, 2509 JG,
The Hague, The Netherlands

Dick G. Simons

Delft University of Technology, Department of Earth Observation and Space Systems, Kluyverweg 1,
2629 HS Delft, The Netherlands

Echo energies of single-beam echo sounders are inverted for the sediment mean grain size via a combination of theoretical and empirical relationships. *In situ* measurements of the seafloor mass density have revealed the presence of a thin transition layer between the water and the sediment. Within this layer, which has a thickness of order 1 cm, the density continuously changes from the water value to the sediment bulk value. The associated impedance gradient affects the normal-incidence reflection coefficient at high frequencies, when the product of wave number and layer thickness is of order unity or higher. A mapping algorithm recognizing this gradient is applied to echo sounder data acquired in three different areas, and for five sounder frequencies between 12 and 200 kHz. Compared with a scheme that relies on the Rayleigh reflection coefficient of a discrete interface, an overall improvement of several phi units in the grain size mapping is achieved by taking the gradient into account. A necessary condition to reach agreement between the acoustic and the ground truth grain size is that the thickness of the transition layer increases with a decreasing grain size. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2345907]

I. INTRODUCTION

The most reliable method to obtain information on the ocean sediment grain size is the gathering of bottom samples followed by a laboratory grain size analysis. However, this process is time consuming and expensive. Remote sensing by acoustic means has long been recognized as a potential method to speed up the process, but there are many problems to overcome. Difficulties include the challenge to separate effects of grain size from those due to density gradients, morphological features, the presence of benthic flora and fauna, anthropogenic influences, etc., which can all influence the process of echo formation.

Acoustic seafloor mapping is an active field of research, which encompasses a variety of sonar systems and processing techniques. Of particular interest to the present study is recent work by Davis *et al.*,[1] who inferred the sediment mean grain size for a Clyde Sea survey area using empirical relationships between sediment properties and the acoustic reflection coefficient. A broadband sonar with a dominant frequency of 1 kHz was used in that work. A similar route is followed in the present paper, but because higher frequencies are used it is found necessary to allow for gradients within the sediment surface layer. Indeed, physical properties such as the porosity and density are often subject to a gradient along the top few centimeters of the sediment.[2–5] The presence of an acoustic impedance gradient considerably influences the normal-incidence reflection coefficient R. At low frequencies the gradient is of little consequence, whereas at

higher frequencies it has the effect of lowering R to below the Rayleigh reflection coefficient of an abrupt interface.[3] This frequency dependence becomes manifest if the product of wave number and transition layer thickness is of order unity or higher.[6] Ainslie[7] summarizes measurements and shows that a more precise condition for the onset of frequency-dependent effects is for the product of acoustic frequency and layer thickness to exceed about 200 m/s. For example, Lyons and Orsi[4] suggest that frequency-dependent measurements[8,9] of normal-incidence reflection loss from a sand sediment may be caused by a strong density gradient at the water-sediment interface. Knowledge of the thickness of the transition layer is required to account for the impedance gradient in modeling the reflection coefficient. In order to explain the measurements of Chotiros *et al.*,[9] the required layer thickness is approximately 20 mm.[7] In the present paper, an empirical relationship based on published data is derived for the thickness of this layer as a function of sediment mean grain size. When high-frequency echo sounder data are inverted for sediment properties, we hypothesize that allowance for the transition layer may yield a substantial improvement of the grain size mapping accuracy. The chief purpose of this paper is to investigate this hypothesis.

It is important that empirical relationships between sediment physical properties and acoustic quantities are used with care, because empirical laws are always based on a limited amount of available data. The nature of any scatter in these data is lost after regression. Data collected in a different part of the world, or in another season, for example, may not be adequately represented by empirical laws. When ground truth is available, the usefulness of empirical relationships can be judged by the extent to which the acoustical

---
[a)]Author to whom correspondence should be addressed; electronic mail: paul.vanwalree@tno.nl

prediction of the grain size matches the ground truth. The present paper evaluates echo sounder data at five frequencies in the range 12–200 kHz, collected in three different geological areas. Ground truth is available for each area in the form of a bottom grab grain size analysis. When acoustic grain size predictions obtained with an impedance step change model and an impedance gradient are compared, the overall improvement offered by gradient is sizable. The degree of the improvement is found to depend on the sounder frequency and geological area, but the observed trend does suggest a wide applicability of the proposed grain size mapping model.

This paper is organized as follows. Section II describes the data processing and the conversion of measured reflection coefficients to sediment mean grain size via empirical relationships. The description starts out with the assumption of an impedance step change, and thenceforth expands the theory to accommodate an impedance gradient. Section III provides an overview of the survey areas, acoustic data, and available ground truth. In Sec. IV the methods described in Sec. II are applied to the data described in Sec. III and the resulting acoustic grain size is presented in maps, and quantitatively compared with the ground truth. Section V summarizes the findings.

II. DESCRIPTION OF THE MAPPING TECHNIQUE

The proposed mapping technique relies on empirical relationships between the sediment grain size and its acoustic properties. Ground truth is available for all test sites described in Sec. III. A subset of the available bottom grabs is used for calibration purposes, and the remaining grabs are used for a comparison between the acoustic prediction of the mean grain size and the ground truth. A straightforward implementation of the mapping method uses the Rayleigh reflection coefficient in conjunction with empirical relationships derived by Bachman[10] relating the sediment grain size to density and sound speed. The Rayleigh reflection coefficient treats the water and the seafloor as two homogeneous media with an abrupt interface, and is independent of the frequency. A more sophisticated approach considers a continuous transition of the impedance, which introduces a frequency-dependent reflection coefficient. Bachman's correlations control the sediment properties beneath the interface. The two approaches are detailed in the following, and grain size maps produced with both methods are presented in Sec. IV. The treatment starts with a complete description of the working method for the discrete interface, and continues with the changes required for the frequency-dependent approach.

A. Reflection coefficient (measurement)

The following description of the working method applies to the data from all test sites. Echo signals are extracted from the recordings as a first step in the data processing. Their envelopes are squared and integrated to yield the echo energy. The integration offers the significant advantage that it does not require algorithms to separate the initial bottom return, associated with the specular reflection from the center of the sonar footprint, from the trailing scattering contributions. Moreover, echo-to-echo energy fluctuations tend to be smaller than echo-to-echo peak amplitude fluctuations. However, there is a built-in assumption that the scattering contributions to the energy obey the same dependence on sediment composition as the specular reflection from a smooth surface of the same impedance. A simple energy conservation argument suggests that this is the case, and a numerical justification is provided by Williams.[11]

The energy flux density E_{RX} of the received echoes depends on the energy E_{TX} of the transmitted sounder pulses according to

$$E_{RX} = \frac{\exp(-4\alpha H)}{4H^2} R^2 E_{TX}, \tag{1}$$

where H denotes the distance between the echo sounder and the seafloor and R the amplitude reflection coefficient of the smooth surface. To discriminate between the energy loss due to transmission into the sediment and attenuation associated with the traveled distance $2H$, the energies are compensated for the spherical spreading factor $1/4H^2$ and the absorption $\exp(-4\alpha H)$. H is determined from the echo return time and the sound speed, and the absorption coefficient α is calculated with the formulas of Francois and Garrison.[12] Temperature and salinity are necessary ingredients in their description of the absorption. *In situ* measurements of the temperature, and a default salinity of 35 psu were used. Remaining uncertainties in the temperature and salinity only have a small effect on the outcome of the calculations, since the water depth difference between the shallowest and deepest parts of each survey area is relatively small.

As the echo sounder source levels are not known and the recorded data uncalibrated, the calculated echo energies have only a relative significance. Absolute values of the reflection coefficient are derived with the help of a few selected bottom grabs that serve to calibrate the energies for an entire data set.

B. Reflection coefficient (model)

1. Impedance step change

To convert measured echo energies into sediment mean grain size,[13] Bachman's regression equations are adopted for the continental shelf and slope environments, denoted by T (terrace) in his paper. That is to say,[10]

$$c_s = 1952 - 86.3M_z + 4.14M_z^2 \tag{2}$$

and

$$\rho_s = 2380 - 172.5M_z + 6.89M_z^2 \tag{3}$$

for the sediment sound speed c_s in m s^{-1}, and the sediment mass density ρ_s in kg m^{-3}, as a function of the mean grain size M_z. The link between the mean grain size and the echo energy is established via the Rayleigh reflection coefficient

$$R = \frac{Z_s - Z_w}{Z_s + Z_w}, \tag{4}$$

where $Z = \rho c$ denotes the characteristic acoustic impedance. For the water side of the interface, the values $\rho_w = 1024$ kg m^{-3} and $c_w = 1529$ m/s are inserted.[14,15] These numbers correspond to Bachman's "standard seawater" with a reference temperature of 23 °C and a presumed salinity of 35 psu. The straightforward equations (1)–(4) suffice to convert echo energies into mean grain size by equating the acoustic reflection coefficient in Eq. (1) to the model reflection coefficient (4), provided that a calibration factor is available. Since the energy of the received echoes E_{RX} is subject to an arbitrary scaling factor, a calibration factor C is required to calculate the R according to

$$R = \frac{2CH}{\exp(-2\alpha H)} \sqrt{E_{RX}}. \tag{5}$$

The factor $C = E_{TX}^{-1/2}$ is a constant for each data set and takes care of dimensional issues. A subset of N bottom grabs is selected for calibration. All echoes within a given search radius from the grab positions are collected and their mean energy is computed. Subsequently, N factors C_n are computed by matching the acoustic reflection coefficient Eq. (5) to the ground truth reflection coefficient Eq. (4), which is calculated via Eqs. (2) and (3) with the grab sample mean grain size. The overall calibration factor C is taken as the root mean square value of the C_n.

A large search radius yields a large number of echoes for which echo-to-echo intensity fluctuations cancel out, whereas a small radius increases the probability that the bottom grab, collected in the center of the search disk, is indeed representative of the echoes. A compromise is found in a value of 200 m, which is used for all data sets described in Sec. III. Typical echo counts for this search disk radius vary from 4 for the Vestfjorden 38 kHz excursion, to 1000 for the Cleaver Bank 12, 38, and 200 kHz survey. The need to perform some averaging is indicated by the standard deviation of the echo energy, which is typically between 30% and 100% of the mean value. Following calibration, a reflection coefficient is obtained for all echoes via Eq. (5). Subsequently, Eqs. (2)–(4) are inverted to find the mean grain size for the individual echoes.

2. Impedance gradient

The above-noted recipe assumes a step change in impedance at the boundary between water and sediment, with uniform properties on either side of the boundary. Thus $R(M_z)$ is independent of frequency. In reality there is always a thin transition region of thickness typically 5–50 mm,[2–5,16–18] implying an average density gradient of order 10^4–10^5 kg m^{-4} (0.01–0.1 g/cm^3/mm). At frequencies of order 100 kHz or higher this gradient cannot be ignored.[3,4,7,19] Specifically, the reflection coefficient is expected to decrease with increasing frequency, owing to the decreasing impedance contrast on a wavelength scale.[3]
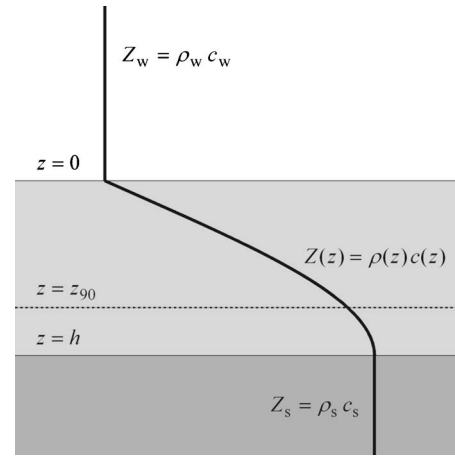
FIG. 1. Illustration of the impedance gradient, Eq. (6), that connects the water impedance Z_w to the bulk sediment impedance Z_s.

In order to calculate the reflection coefficient, we assume that uniform semi-infinite layers of impedance Z_w (water) and Z_s (sediment) are separated by a continuous transition layer of finite thickness h (Fig. 1). The transition layer is modeled as an impedance profile of the form

$$Z(z) = Z_w \frac{\text{sech}^2(az/h)}{[1 - \tanh(a)\tanh(az/h)]^2}, \tag{6}$$

$$a = \cosh^{-1} \sqrt{\frac{Z_s}{Z_w}}. \tag{7}$$

This functional form ensures a continuous impedance at both boundaries and a continuous gradient dZ/dz at the lower boundary. We further assume that the normal-incidence reflection coefficient is a function of the impedance profile alone and does not depend on density and sound speed separately. With this second assumption we can simplify the calculations by constructing an effective density profile

$$\rho_{eff}(z) = \frac{\rho(z)c(z)}{c_w} \tag{8}$$

such that, with $c_{eff}(z) = c_w$ everywhere, the impedance is identical to that of the equation for $Z(z)$. The advantage of this specific functional form is that it is amenable to analytical solution.[6,20] Substrate properties (ρ_s, c_s) are from Bachman.[10] The sediment absorption coefficient is unimportant for this calculation, and is chosen arbitrarily to be zero.

It is necessary to obtain an estimate of the transition layer thickness before the above-presented theory can be applied to our echo sounder data. For this purpose the quantity z_{90} is introduced, which denotes the depth in the transition layer at which 90% of the density difference between the water and the sediment is reached. Figure 2 shows z_{90} values, inferred from measured density or porosity profiles,[3–5,17,21,22] for different grain sizes ranging from 1-2 φ (medium sand) to 9-10 φ (medium clay). (φ is defined as $-\log_2$ of the grain diameter in millimeters) A least-squares regression is carried out on these data (shown as a solid line in the figure), resulting in the empirical fit

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

van Walree et al.: Mean grain size mapping   2557
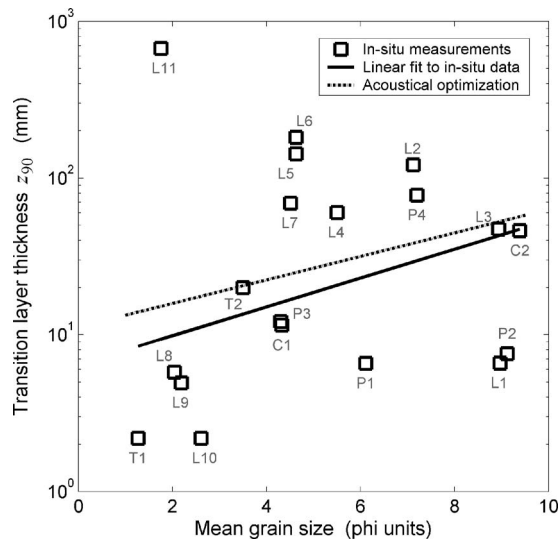
FIG. 2. *In situ* measurements of the layer thickness parameter z_{90} vs grain size. The origin of these data is as follows. Cn: Carbó (Ref. 3) with C1 = sand, C2 = mud. Ln: Lyons and Orsi (Ref. 4) with L1=227-BS-BC, L2=260-BS-BC, L3=264-BS-BC, L4=268-BS-BC, L5=89-1-KW-DC, L6 =89-2-KW-DC, L7=124-1-KW-DC, L8=413-2-PC-DC, L9=413-3-PC-DC, L10=490-PC-DC, L11=250-BS-BC. Pn: Pouliquen and Lyons (Ref. 5) with P1=Venere Azzurra, P2=Portovenere, P3=Tellaro, P4=Punta della Mariella. T1: Tang *et al.* (Ref. 17) T2: Tang (Ref. 21) for the z_{90} value and Miller (Ref. 22) for the corresponding grain size value. The solid line is a linear regression to these measurements, with the outlier marked L11 excluded: $\log_{10}z_{90}=0.092\times M_z+0.81$. The dashed line results from the acoustical optimization described in Sec. IV C: $\log_{10}z_{90}=0.075\times M_z+1.05$.

$$\log_{10}z_{90} = (0.092 \pm 0.052) \times M_z$$
$$+ (0.81 \pm 0.31) \quad (z_{90} \text{ in mm}). \quad (9)$$

The layer thickness h follows from the functional form Eq. (6) as $h=z_{90}/0.652$. Two points from the Lyons and Orsi measurements[4] are omitted from the regression. The first one is labeled L11 in Fig. 2 and is left out because it is an outlier whose inclusion strongly influences the outcome of the regression. The second omission is a measurement with a very thin transition layer, resulting in a zero value of the transition layer thickness z_{90}. All other available *in situ* measurements in Fig. 2 are used for the regression.

The value of the data plotted in Fig. 2 could be increased by including error estimates for the z_{90} measurements. Such estimates could be made from the original measurements, but we are not able to estimate the uncertainty in layer thickness from the publications cited. In the case of one publication,[5] it is likely that the method used to analyze the profiles results in a lower limit for the z_{90} value.

Although the observed correlation in Fig. 2 is not overwhelming, its use at least allows the incorporation of first-order frequency-dependent effects. The validity of this assumption can be judged by the success or otherwise of the overall method in the inversion of the reflection coefficient —which in turn can be judged by comparison with grain size analysis of the grab samples.

With the above-presented assumptions and approximations, the reflection coefficient R is uniquely determined by the frequency f and mean grain size M_z. The reflection coefficient is calculated using the method of Ainslie.[20] Figure 3
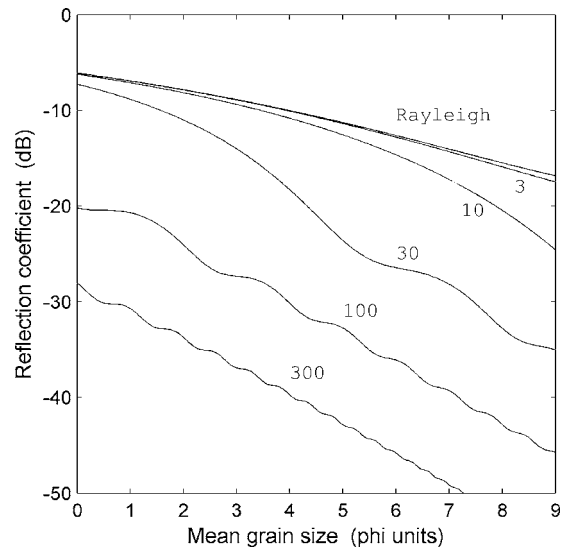
FIG. 3. The reflection coefficient $R(M_z)$ for a set of frequencies between 3 and 300 kHz. In the low-frequency limit the appropriate curve, marked "Rayleigh," is equivalent to the frequency independent reflection coefficient of Sec. II B.

shows the reflection $20\log_{10}|R|$ vs M_z for a pertinent set of frequencies. For any known frequency a curve such as those in Fig. 3 can be used as a lookup table to invert R measurements (if calibrated) for the mean grain size. Results presented in Sec. IV are calculated in this manner.

III. OVERVIEW OF THE SURVEY AREAS

The echo sounder data under consideration were acquired during five separate sea trials covering three different areas. Sediment samples were collected in all areas for ground truth. One particular area, the Cleaver Bank, has been surveyed on three occasions, with a total of five different sounder frequencies. In the following the five acoustic data sets and the ground truthing procedures are briefly described.

A. Cleaver Bank

The Cleaver Bank, some 130 km northwest of the Netherlands (Fig. 4), was surveyed with a 150 kHz echo sounder in October 2000. It is a part of the North Sea that offers a sedimentary diversity from soft and flat ("sandy Mud") to hard and rough ("sandy Gravel"). The ship sailed ten longitudinal tracks on a square measuring 10×10 nautical miles. The water depth in the area varies between 30 and 60 m. Raw echo sounder signals were recorded for offline analysis. Soon after, in November 2000, the Cleaver Bank was revisited with a 66 kHz sounder. The same ten legs were sailed. Finally, in November 2004, the area was subject to a multibeam echo sounder expedition. In this last survey, a dense pattern of east-west tracks was sailed (shown in the inset of Fig. 4). A triple-frequency, single-beam echo sounder operating at 12, 38, and 200 kHz was also operated during this survey, which yielded valuable echo sounder data at three new frequencies.

A comprehensive description of the Cleaver Bank geology and ground truth is found in Van Walree *et al.*[23] Only a summary of the ground truthing is given here. Fifty grab
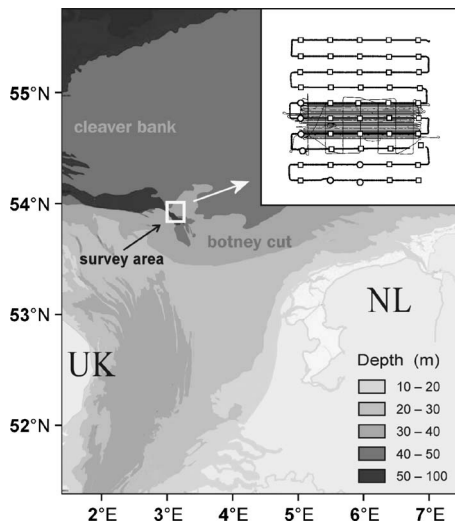
FIG. 4. Overview map of the Cleaver Bank trials area between United Kingdom and the Netherlands. The inset shows a magnified portion with tracks sailed on several occasions as well as the October 2000 bottom grabs (white markers).

FIG. 5. Map of the Norwegian Vestfjorden trials area. The echo sounder survey consists of two isolated grids and features a number of bottom grabs indicated by white markers.

samples were collected with Van Veen and Hamon grabs during the original October 2000 sea trials. The laboratory analysis of the bottom grabs consisted of the following steps. The samples were dried and sieved with a 2 mm mesh to separate the gravel and shells from the smaller grains. Subsequently the biogenic and clastic components were separated to determine the shell and gravel fractions of each grab. The remaining sand and mud portions of the original grab were subjected to a granulometric analysis by optical microscopy, after which the respective sand and mud fractions could be determined. To remove a gravel bias from the mean grain size, the cumulative grain size distribution was rescaled to a value of 100% at a grain size diameter of 2 mm. Subsequently, the mean grain size M_z was calculated according to

$$M_z = \frac{3\varphi_{10} + 4\varphi_{50} + 3\varphi_{90}}{10}, \tag{10}$$

where the phi values are defined by minus $\log_2$ of the grain diameter in millimeters. The subscripts indicate percentiles such that, for example, 10% of the grab has a grain size smaller than φ_{10}. The definition, Eq. (10), departs from the more common definition, Eq. (11), because the laboratory analysis returned the φ_{10} and φ_{90} percentiles rather than the φ_{16} and φ_{84} percentiles.

The 4-year time span between the grab sample analysis in 2000 and the survey in 2004 introduces some uncertainty in the validity of the ground truth. Previous work[23] revealed that the ground truth obtained in 2000 showed a great overall correspondence with a geological map from 1987. Therefore one may presume that overall seabed changes over a time span of four years are small in this area. Nonetheless, caution is required with local assessments of the sediment type, for instance because of the presence of wandering sand ribbons that overlie the gravel bed.[23]
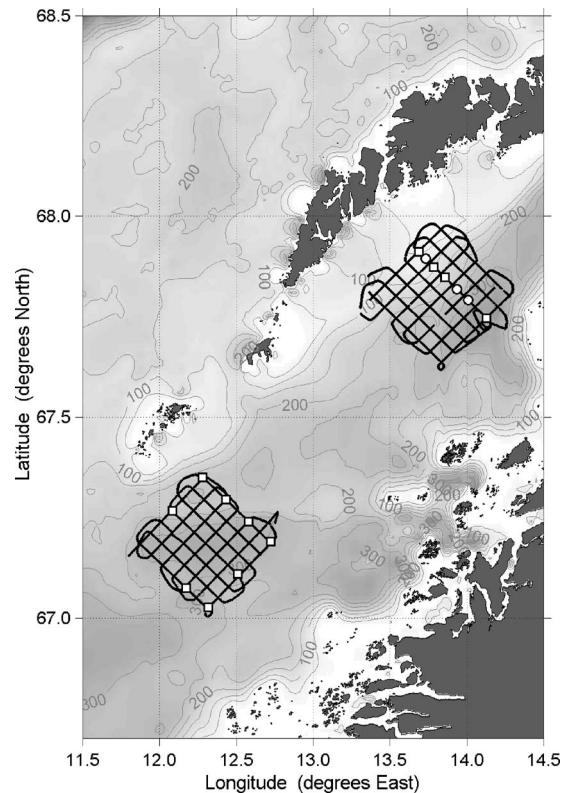
B. Vestfjorden

The Norwegian Vestfjorden area was surveyed with a 38 kHz echo sounder in May/June 2001. Two sites along the axis of the fjord were selected for the acoustic measurements, each with dimensions of roughly 15×15 nautical miles. The water depth ranges from 100 to 350 m. At both sites the ship track forms a grid of seven by seven legs in the shape of a carpet beater (Fig. 5). Fifteen bottom grabs were collected, spread over the two sites, whose laboratory analysis is detailed in Ref. 24. Incidental gravel was left out of the cumulative distribution. The mean grain size was calculated according to Folk,[13]

$$M_z = \frac{\varphi_{16} + \varphi_{50} + \varphi_{84}}{3}. \tag{11}$$

C. Stavanger

The final test site is located in the coastal area of Norway, in the vicinity of the city of Stavanger (Fig. 6). Measuring approximately 20×20 nautical miles, a grid was surveyed with a 38 kHz sounder in October 2002. The bathymetry is relatively flat with a depth between 250 and 285 m (Fig. 10), except for a few sea-mounts in the eastern part of the area, near the island of Utsira, which rise some 50 to 100 m above the nominal seabed. Thirty-five bottom grabs were collected along one horizontal and one vertical leg of the survey. The laboratory analysis is described by Ref. 25.
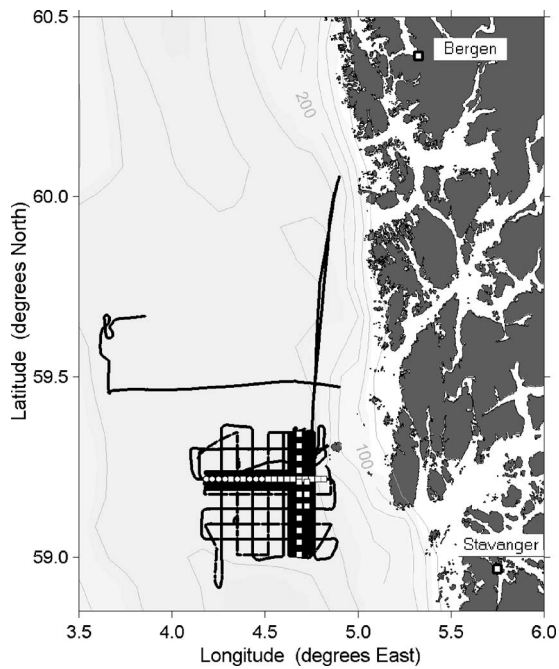
FIG. 6. Map of the Norwegian trials area near Stavanger. The echo sounder survey consists of a sparse grid, with a more dense series of legs surrounding the ground truth regions indicated by the white markers.

Pebbles were removed from the grab before the grain size distribution was determined. The mean grain size was subsequently calculated via Eq. (11).

IV. RESULTS AND DISCUSSION

A. Grain size maps

The calculations presented in Sec. II lead to a recipe for converting echo energies into mean grain sizes. Some averaging is required in order to map the mean grain size without the distraction of echo-to-echo fluctuations. The most convenient averaging method is found to depend on the data set under examination. For all cases of sailed grids with a large separation between the legs, a moving median filter of length 25 is applied to the echo energies before the mean grain size is inferred. A median filter removes outliers, while it preserves step changes of the input parameter. This procedure is applied to all but the Cleaver Bank 2004 data with the 12, 38, and 200 kHz sounders. For the latter survey the legs are closely spaced, and the number of echoes (7×10^5 at each frequency) is large enough to support a map with full seafloor coverage. To this end the relevant section of the seafloor is divided into 490×45 bins measuring 33×110 m. For each bin the mean echo energy is calculated, and this value is turned into a mean grain size value for that particular bin by the methods of Sec. II.

In the following the mean grain size mapping is qualitatively illustrated with three example maps. Subsequently the fidelity of the mapping algorithms is quantified for the entire collection of echo sounder data by means of scatter plots. Figure 7 presents the results for the 150 kHz Cleaver Bank sea trials conducted in October 2000. The first graph gives the bathymetry and the other graphs show the acoustic mean grain size obtained on the assumption of an impedance step
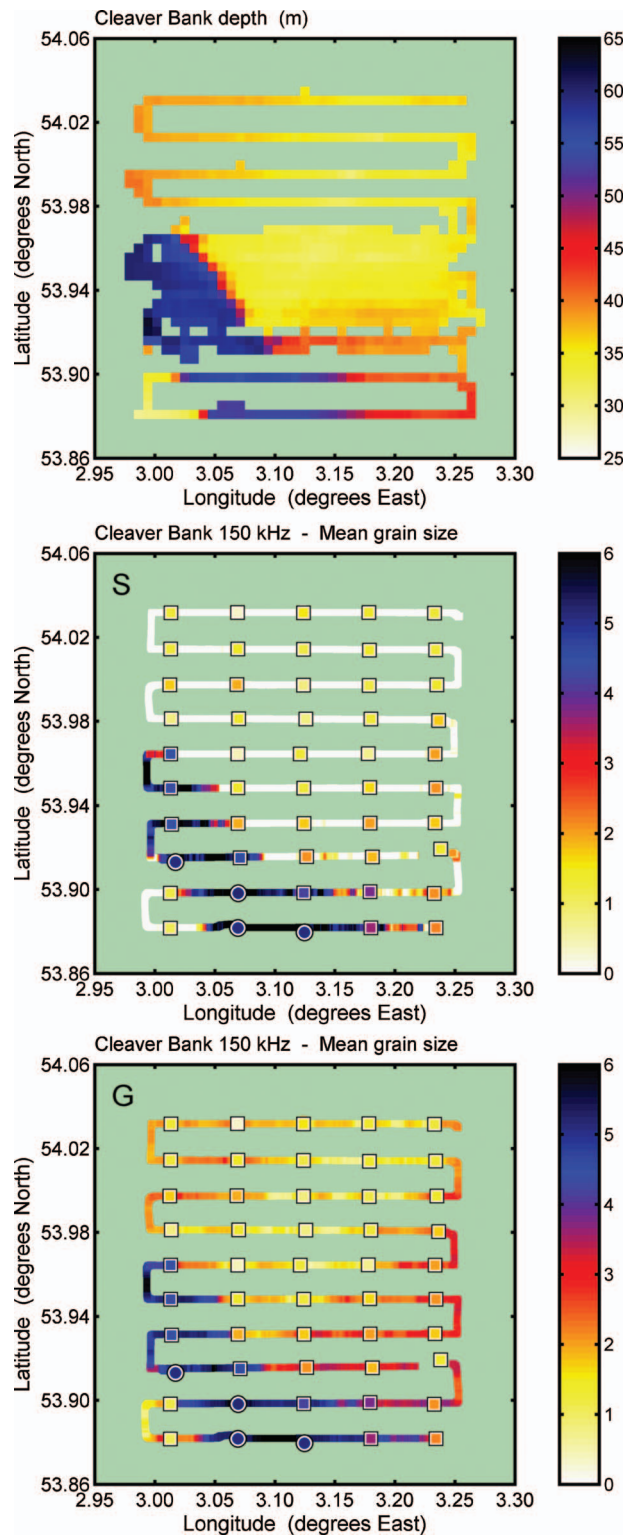
FIG. 7. Top graph: bathymetry of the Cleaver Bank survey area. Middle graph: 150 kHz mean grain size map obtained with the impedance step change (S). Bottom graph: alternative grain size map obtained with the impedance gradient (G).

change (S) or gradient (G). The ground truth mean grain size is also included. A round marker indicates that a grab is used for the calibration procedure. Thus, on average the acoustic grain size is forced to match the ground truth for the round markers. The merits of the mapping technique should therefore be judged from the regions adjacent to the remaining
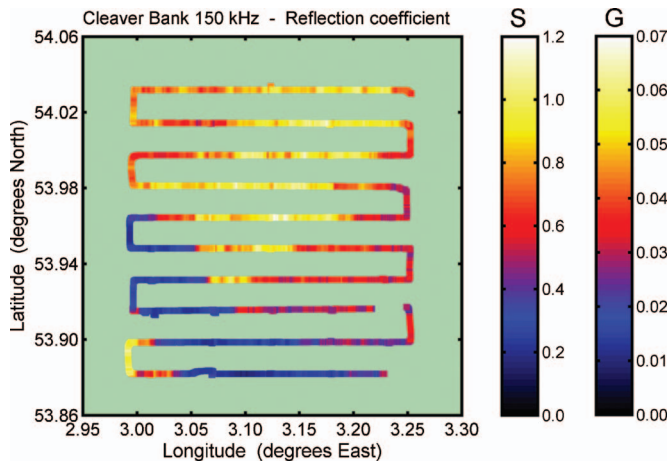
FIG. 8. Amplitude reflection coefficient for the Cleaver Bank 150 kHz data. The maps for the impedance step change (*S*) and impedance gradient (*G*) differ by a scaling factor only.

square markers, which are not used for calibration. Application of the abrupt impedance transition leads to a considerable mismatch for the great majority of square markers, which represent the sandy and gravelly parts of the survey area. The dynamic range of the acoustic mean grain size is simply too large. Matters are noticeably improved by replacing the discrete impedance mismatch with a gradient. The dynamic range of the acoustic mean grain size is decreased and the resulting map resembles the ground truth with a remarkable overall fidelity.

For the sake of completeness an intermediate step in the calculation, the amplitude reflection coefficient, is presented

for the Cleaver Bank 150 kHz data. It is plotted in Fig. 8 for both impedance transition types. In case of the step change the reflection coefficient is unrealistically high and the physically impossible $R > 1$ occurs. In contrast, the impedance gradient lowers the reflection coefficient by an order of magnitude, consistent with the measurements of Kimura and Tsurumi[26] or the predictions of, e.g., Carbó[3] and Lyons and Orsi.[4] The *S* and *G* reflection coefficients in Fig. 8 are identical apart from a scaling factor. Indeed, the mean grain size calibration procedure described in Sec. II just leads to a different echo energy scaling for the *S* and *G* cases. This applies not only to the Cleaver Bank 150 kHz data but to all sounder frequencies and survey areas. Maps of the reflection coefficient are omitted for the remaining examples.

The second example is the Cleaver Bank 38 kHz data set, illustrated with mean grain size maps in Fig. 9. The change from the *S* to the *G* impedance profile lowers the contrast of the acoustic mean grain size, similar to the 150 kHz case. However, where the modification was about right at 150 kHz, the gradient overcompensates at 38 kHz. On the chosen color scale the result is close to a red-blue binary grain size. Again the mud trench at the left side is tuned to the proper grain size value by the calibration procedure. The sandy part of the survey area (bottom right) roughly achieves the same blue grain size as the mud calibration area. In contrast to the blue, there is a red colored area representing the gravelly sand and sandy gravel described in Ref. 23. Whereas the acoustic grain size exceeds the ground truth values for the *S* transition, the situation is reversed when the gradient is applied.
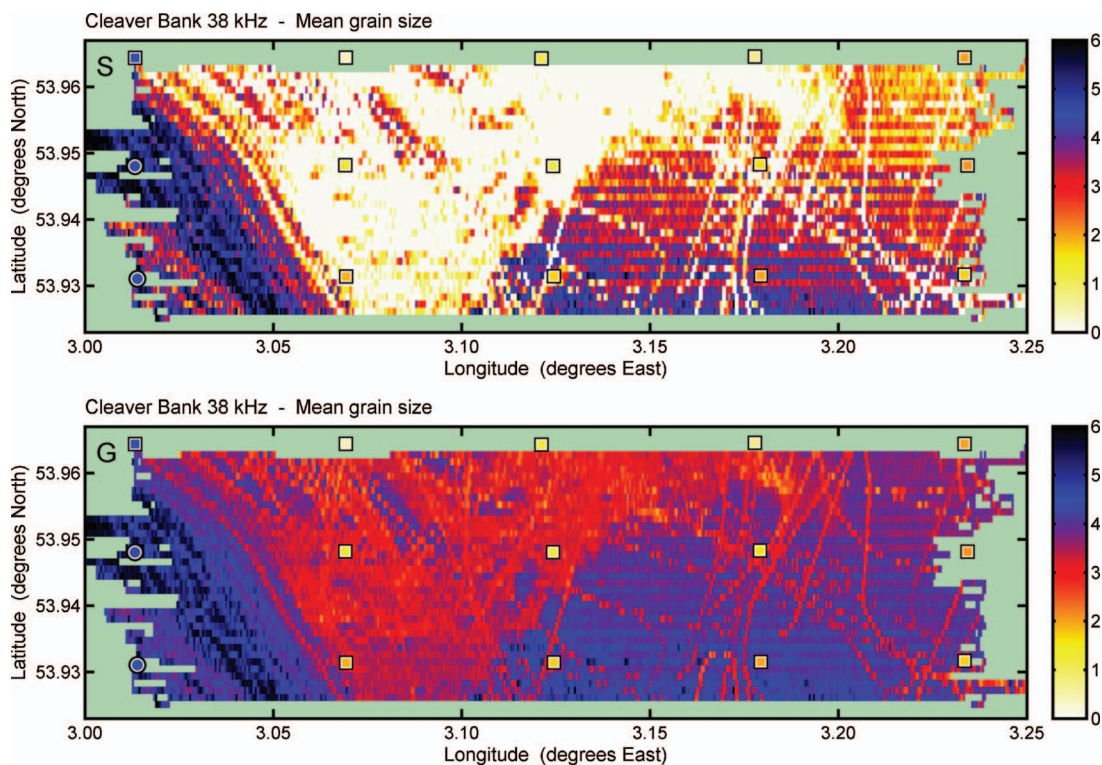
FIG. 9. Cleaver Bank 38 kHz grain size maps obtained with the *S* (top) and *G* (bottom) impedance transitions. The horizontal zebra stripe pattern noticeable in the right half is an artifact apparently related to the ship sailing direction. Calibration grabs are indicated by round markers and the remaining grabs by square markers.

There are two additional features in Fig. 9 that leap to the eye. First, the right half of each map is characterized by a pattern of horizontal stripes. This phenomenon appears to be related to the ship sailing direction, alternately eastward and westward, although the mechanism remains unclear. The dependence of the echo energy on the ship course is also present at 200 kHz, and to a lesser degree at 12 kHz. A perhaps more striking feature of Fig. 9 is the occurrence of a large number of thin-linear features, with a marked deviation of ~2 phi units in the inferred mean grain size, corresponding to higher echo energies. The features run in arbitrary directions and are independent of the ship sailing direction. They are also clearly visible in the 200 kHz data, whereas they are just barely noticeable in the 12 kHz data. Section IV D discusses the possible origin of these thin-linear features and their relevance.

As a third and final example the mean grain size is mapped for the Stavanger 38 kHz data in Fig. 10. This trials area has an overall finer-grained seafloor than the Cleaver Bank and features a five- to tenfold increase of the water depth. The bathymetry in the top graph is relatively flat, and, unlike the Cleaver Bank area, there is no strong correlation between the depth and the sediment type. An exception is found in the neighborhood of the bathymetrical features (sea mounts) toward the Norwegian coast, at the east side of the grid. This is where the bottom grabs with the largest grains were collected. The calibration grabs were chosen on a section of the horizontal leg where both the grab sample constitution and the echo energies vary little. Two grabs in this region contained some gravel and were left out of the calibration set.

The mean grain size map obtained by application of the impedance step change is shown in the middle panel. Again, the contrast of the acoustic grain size again is too high, similar to the Cleaver Bank results in Figs. 7 and 9. Use of the impedance gradient yields a noticeably better agreement between the acoustic prediction and the ground truth.

B. Scatter plots

So far the merits of the method proposed in Sec. II have been judged by visual comparison of acoustic grain size maps with the ground truth. For a more quantitative comparison all available data are combined into scatter plots (Fig. 11), including the Vestfjorden area and the sounder frequencies not covered by Figs. 7, 9, and 10. Figure 11 shows the acoustic versus the ground truth mean grain size, calculated for the S and G impedance transition types. Survey areas and echo sounder frequencies are differentiated by markers with various shapes and colors. The acoustic grain size was calculated from the arithmetic mean energy of all echoes within a search radius of 200 m from each grab. Notice that the area of the markers in Fig. 11 is proportional to the number of echoes found. Tiny markers carry less significance than big ones.

It is recalled that there are calibration grabs that are forced to coincide, on average, with the ground truth. These calibration grabs are among the softer sediments for each data set. The performance of the algorithms must therefore
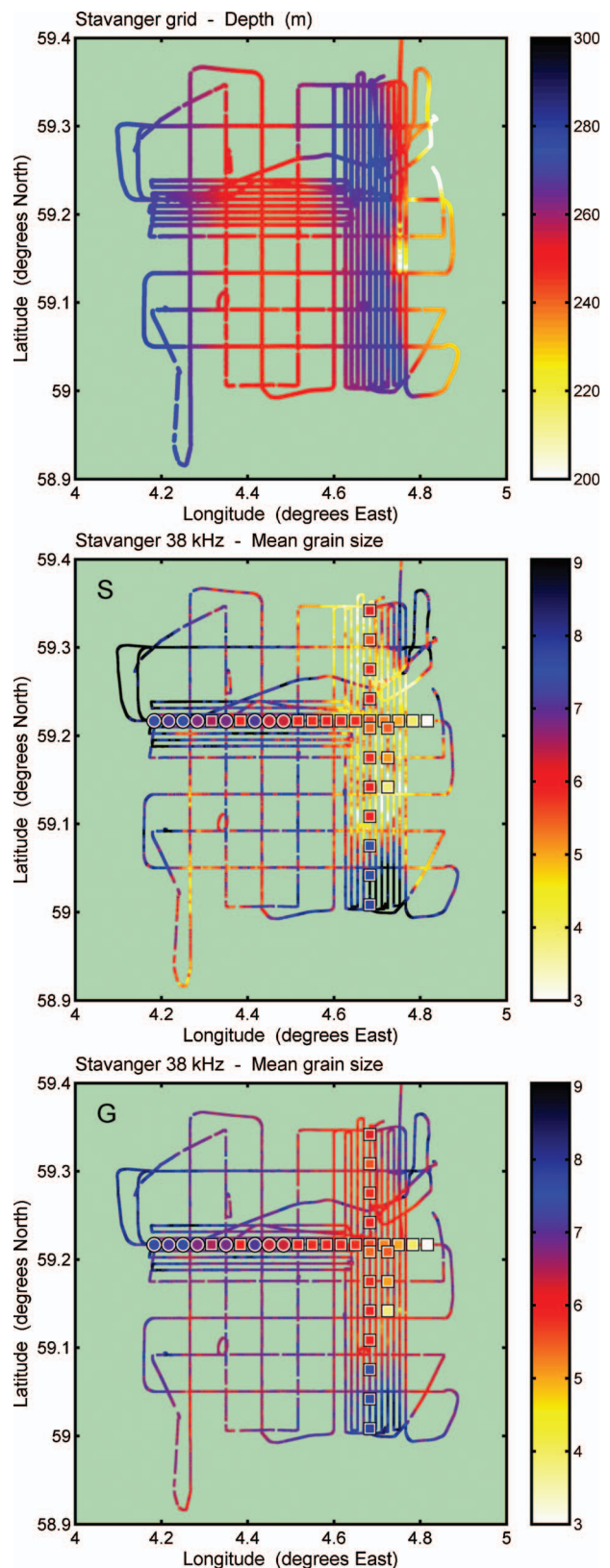
FIG. 10. Bathymetry of the Stavanger 38 kHz survey (top graph). Mean grain size maps obtained with the impedance step change (middle graph) and impedance gradient (bottom graph) are shown. Again, calibration grabs are indicated by round markers and the remaining grabs by square markers.

follow from inspection of the larger grains toward the left side of Fig. 11. On the whole the impedance step change yields a considerable amount of scatter and an overall depar-
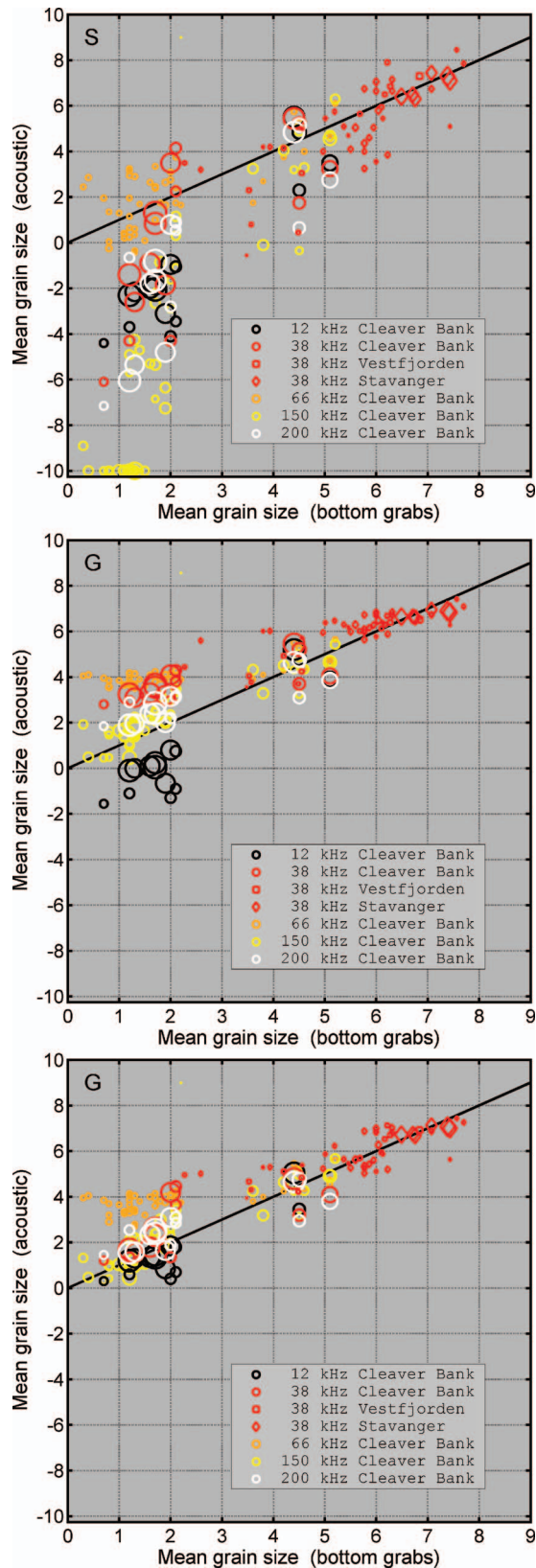
FIG. 11. Scatter plots that combine the results for all survey areas and frequencies. Results are shown for the S and two G impedances. The middle panel corresponds to a gradient given by the fit to the *in situ* layer thickness measurements (solid line in Fig. 2), whereas the bottom panel shows the minimum rms case (dashed line in Fig. 2; explained in Sec. IV C). Marker areas are proportional to the number of echoes used for the acoustic grain size calculation. Notice that the acoustic grain size has been clipped to a value of −10 for the topmost graph.

TABLE I. Correlation coefficients and weighted rms values for all survey areas and sounder frequencies; grabs denotes the number of contributing grabs.

| Survey | Grabs | Impedance step change | | Impedance gradient | |
		Correlation coefficient	rms	Correlation coefficient	rms
Cleaver Bank 12 kHz	16	0.92	3.52	0.92	1.63
Cleaver Bank 38 kHz	16	0.71	2.65	0.73	1.68
Cleaver Bank 66 kHz	48	0.74	1.13	0.77	2.32
Cleaver Bank 150 kHz	48	0.71	15.7	0.75	0.67
Cleaver Bank 200 kHz	16	0.80	4.29	0.82	0.87
Vestfjorden 38 kHz	14	0.79	1.51	0.85	0.68
Stavanger 38 kHz	34	0.69	0.81	0.71	0.66
All together	192	0.59	7.91	0.77	1.36

ture from the ideal solid line $M_{z,\text{ac}}=M_{z,\text{bg}}$, where $M_{z,\text{ac}}$ denotes the acoustic mean grain size and $M_{z,\text{bg}}$ the ground truth (bottom grab) mean grain size. The error is largest for the Cleaver Bank 150 kHz data, followed by the 200 kHz frequency. At 66 kHz the $M_{z,\text{ac}}$ seem about right, albeit with some scatter around the mean. For the finer-grained Norwegian test sites ($f=38$ kHz) the overall mismatch is smaller, but still the deviations range up to 4 phi units.

A very different picture is obtained for the impedance gradient in the middle graph. At each frequency the amount of scatter is decreased, and on the whole the acoustic mean grain size is closer to the ground truth. The scatter reduction is most obvious at 150 kHz, where the impedance gradient brings the acoustic mean grain size remarkably close to ground truth. Substantial progress is also made at 200 kHz. The Cleaver Bank 12 kHz data remain below the ideal line, whereas the Cleaver Bank 38 and 66 kHz data end up above this line. Within each data set the impedance gradient reduces the scatter along the ordinate, but the slope is subject to a residual error. Figure 11 also shows that the 38 kHz data, originating from three different areas, form a consistent subset of the scatter plot without apparent discontinuities between test sites.

Two quantities are calculated to quantify the influence of the impedance transition type on the conversion of echo energy to grain size: the correlation coefficient between $M_{z,\text{ac}}$ and $M_{z,\text{bg}}$, and the rms value of their difference, weighted by the number of echoes K_n contributing to the nth grab acoustic grain size

$$\text{rms} = \left(\frac{\sum\limits_{n=1}^{N} K_n (M_{z,\text{ac}}(n) - M_{z,\text{bg}}(n))^2}{\sum\limits_{n=1}^{N} K_n} \right)^{1/2}. \tag{12}$$

The weighting diminishes the contribution of grabs with poor statistics, such as the 150 kHz outlier with $M_{z,\text{ac}}>8$, for which there are only two echoes within the search radius. Nonetheless, the overall influence of the weighting on the rms values is small.

Table I shows that the impedance gradient improves the

correlation coefficient for all data sets. An improvement is also observed when all grabs and sounder frequencies are considered together. The so-called P values associated with the correlation coefficients are not tabulated. P gives the probability that the calculated correlation coefficient arises by chance. Since all P are vanishingly small ($\ll 10^{-2}$) the calculated correlation coefficients are significant.

Despite the strong correlations, the correlation coefficient does not properly reflect the merits of the impedance gradient. The correlation coefficient tells us to what extent there exists a linear relationship between the acoustic and the grab sample mean grain size. It neglects the strength of the relationship, represented by the slope of $M_{z,ac}$ vs $M_{z,bg}$ in the scatter plots. For this reason the correlation coefficient achieves high values already for the discrete impedance interface, despite the considerable mismatch in absolute terms. $M_{z,ac}$ and $M_{z,bg}$ are well correlated for either impedance transition type. The rms values in Table I provide a more useful criterion to judge the benefits of the methods. These values clearly demonstrate the overall superiority of the impedance gradient. For all frequencies and areas the gradient method delivers a substantial improvement, except for the Cleaver Bank 66 kHz data. Although the impedance gradient does reduce the scatter within the latter data set, the $M_{z,ac}$ are actually pushed away from $M_{z,bg}$. In fact, with both the 38 and 66 kHz sounders it is difficult to discriminate between sandy and muddy sediments in the Cleaver Bank area.

C. Dependence of the transition layer thickness on grain size

Sections IV A and IV B showed grain size maps and scatter plots based on a regression of the form $\log_{10} z_{90} = aM_z + b$ [see Eq. (9) and Fig. 2]. There is, however, a sizable scatter in the underlying *in situ* measurements. In order to get an idea of the sensitivity of the mapping fidelity to the values of the slope a and intercept b, the mapping is repeated for several combinations of a and b values. For each combination the steps described in Sec. II are repeated, including calibration, and the total weighted rms is computed for all available data (cf. the bottom row in Table I). Figure 12 shows the outcome of these calculations. The smallest encountered rms values are of order unity, with some ambiguity in the slope and a considerable ambiguity in the intercept. The latter ambiguity is due to the calibration procedure, which, regardless of the nominal layer thickness, forces the acoustic grain size to coincide with that of the calibration grabs. (The corresponding amplitude reflection coefficient drops rapidly with an increasing intercept value and achieves totally unrealistic values of order 10^{-3} at $b = 2.5$.) Figure 12 has a minimum rms of 0.90 at $a = 0.075$ and $b = 1.05$. A scatter plot is created for these values and shown in the bottom graph of Fig. 11. The low rms is evidenced by a further reduction of the scatter, compared with the middle graph that was obtained for the fit to the *in situ* transition layer thickness data. Most markers are now within 1 phi unit of the ideal solid line. Little improvement is nonetheless observed at 66 kHz, which data carry little weight and have a small vote in the total weighted rms.
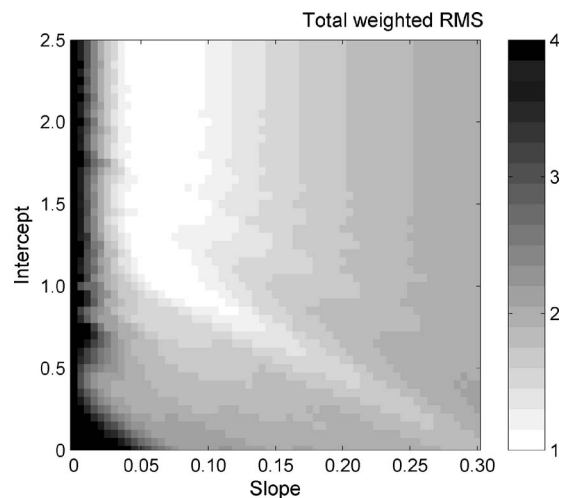
FIG. 12. Total weighted rms difference [Eq. (12)] between the ground truth and the acoustic prediction of the mean grain size. The rms values are calculated for a grid of combinations of the slope a and intercept b of the assumed relationship $\log_{10} z_{90} = aM_z + b$ (z_{90} in millimeters).

Perhaps the most important conclusion from Fig. 12 is that a constant transition layer thickness does not work. A zero (or negative) slope leads to acoustic grain size predictions that disagree strongly with the ground truth. The functional form $\log_{10} z_{90} = aM_z + b$ was just chosen as the simplest assumption of some existing relationship between z_{90} and M_z, and there may be better guesses. Nonetheless, in order to account for the measured echo energies, regardless of the precise functional form it appears that the transition layer thickness must increase with a decreasing grain size.

D. Plough marks

It is hypothesized that the thin-linear features observed in the grain size maps of Fig. 9 correspond to furrows caused by fishing gear dragged over the seafloor. Beam trawlers were witnessed during the November 2004 acoustic survey, and are known to operate regularly in this part of the North Sea. The hypothesis is corroborated by a morphological map (not shown) produced by means of the multibeam echo sounder operated during the same survey. This map shows a crisscross of "plough marks" normally associated with trawling, some of which precisely match the coordinates of the thin-linear features in Fig. 9. A closer examination of the Cleaver Bank 38 kHz data reveals that the echo energy in some furrows rises to as much as 10 dB above that of the surroundings. It is also observed that the excess energy is carried by the initial bottom return and that the energy in the echo tail is not noticeably altered. The energy rise is also pronounced at 200 kHz, whereas it is only weakly present in the 12 kHz data.

Research on the physical effect of beam trawls on the seafloor has shown that the penetration depth in sandy sediments is typically of order 5 cm.[27,28] There is a large spread in reported lifetimes of the plough marks, which may last from less than one day to over one year,[29,30] depending on the environment, type of fishing gear, and the method of monitoring. A record of beam trawl disturbance in another part of the North Sea[31] shows that the surface roughness in a

fishing sector characterized by sand waves and some ripples was reduced, but that the particle size distribution was not altered. Similar conclusions were obtained for an experiment in the Adriatic Sea.[32] A recent cruise in the Bering Sea[33] revealed mean grain size changes of only ~0.02 phi units before and after experimental trawling on a sandy seafloor. The dramatic increase in the echo energy during the Cleaver Bank 2004 survey suggests that the trawl force not only flattens the surface, but that it may also compress the top layer. Alternatively, upwhirling surface particles may be carried away by currents to deposit elsewhere, promoting the more tightly packed subsurface layer to the new top layer. Destruction of the impedance gradient could thus account for the increased acoustic mismatch. Although the explanation of the plough mark acoustics is not central to the present paper, their pronounced acoustic fingerprint at least illustrates the importance of the condition of the sediment top few centimeters on the normal-incidence reflection coefficient.

V. CONCLUSIONS

In previous work, empirical relationships were used to transform measured reflection coefficients of a low-frequency sonar into sediment mean grain size.[1] A central assumption was the applicability of the Rayleigh reflection coefficient, describing a discrete interface between two homogeneous media. In the present paper it is shown that this assumption is no longer valid at higher frequencies of, say, 10 kHz and beyond. A considerable improvement in the grain size conversion is obtained by allowing for the presence of a thin transition layer, within which the acoustic impedance gradually changes from the water value to the sediment bulk value. This impedance gradient is due to a density gradient of order $5 \times 10^4 \, \text{kg m}^{-4}$, which stretches over a layer of only a few centimeters. A grain size mapping algorithm that takes this impedance gradient into account is applied to echo sounder data collected at three sites, and for a total of five sounder frequencies in the 12–200 kHz range. An overall improvement of several phi units is obtained over an algorithm employing the Rayleigh model. Out of seven data sets, the only one not to show an improvement is the Cleaver Bank 66 kHz survey. Residual errors of ~2 phi units between the acoustic and the ground truth grain size, even with the gradient correction, are nonetheless still sizable, which may point to incompleteness of the physical model or to survey areas not adequately described by the employed empirical relationships. It is finally shown that the acoustic grain size prediction is irreconcilable with the ground truth if a constant transition layer thickness is adopted. Agreement requires the layer thickness to increase with a decreasing grain diameter.

ACKNOWLEDGMENTS

[1] A. Davis, R. Haynes, J. Bennell, and D. Huws, "Surficial seabed sediment properties derived from seismic profiler responses," Marine Geology **182**, 209–223 (2002).

[2] D. R. Jackson and K. B. Briggs, "High-frequency bottom backscattering: Roughness versus sediment volume scattering," J. Acoust. Soc. Am. **92**(2), 962–977 (1992).

[3] R. Carbó, "Wave reflection from a transitional layer between the seawater and the bottom," J. Acoust. Soc. Am. **101**, 227–232 (1997).

[4] A. P. Lyons and T. H. Orsi, "The effect of a layer of varying density on high-frequency reflection, forward loss, and backscatter," IEEE J. Ocean. Eng. **23**(4), 411–422 (1998).

[5] E. Pouliquen and A. P. Lyons, "Backscattering from bioturbated sediments at very high frequency," IEEE J. Ocean. Eng. **27**(3), 388–402 (2002).

[6] A. J. Robins, "Reflection of plane acoustic waves from a layer of varying density," J. Acoust. Soc. Am. **87**, 1546–1552 (1990).

[7] M. A. Ainslie, "The effect of centimetre scale impedance layering on the normal-incidence seabed reflection coefficient," *Proceedings of Boundary Influences in High Frequency, Shallow Water Acoustics*, Bath, UK, 2005, pp. 423–430.

[8] N. P. Chotiros, "Inversion and sandy ocean sediments," *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diachock, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht, 1995), pp. 353–358.

[9] N. P. Chotiros, A. P. Lyons, J. Osler, and N. G. Pace, "Normal incidence reflection loss from a sandy sediment," J. Acoust. Soc. Am. **112**(5), 1831–1841 (2002).

[10] R. T. Bachman, "Acoustic and physical property relationships in marine sediment," J. Acoust. Soc. Am. **78**, 616–621 (1985).

[11] K. L. Williams, "Determining flat interface reflection coefficients using forward scattering from a rippled sediment interface," J. Acoust. Soc. Am. **117**, 2434–2435 (2005).

[12] R. E. Francois and G. R. Garrison, "Sound absorption based on ocean measurements. II. Boric acid contribution and equation for total absorption," J. Acoust. Soc. Am. **72**, 1879–1890 (1982).

[13] R. L. Folk, "A review of grain-size parameters," Sedimentology **6**, 73–93 (1966).

[14] A. D. Pierce, *Acoustics, An Introduction to its Physical Principles and Applications*, ed. (Acoustical Society of America, New York, 1994), p. 34.

[15] K. V. Mackenzie, "Nine-term equation for sound speed in the oceans," J. Acoust. Soc. Am. **70**, 807–812 (1981).

[16] T. H. Orsi, A. L. Anderson, and A. P. Lyons, "X-ray tomographic analysis of sediment macrostructure in Eckenförde Bay, western Baltic Sea," Geo-Mar. Lett. **16**, 232–239 (1996).

[17] D. Tang, K. B. Briggs, K. L. Williams, D. R. Jackson, E. I. Thorsos, and D. B. Percival, "Fine-scale volume heterogeneity measurements in sand," IEEE J. Ocean. Eng. **27**(3), 546–560 (2002).

[18] R. A. Wheatcroft, "In situ measurements of near-surface porosity in shallow-water marine sands," IEEE J. Ocean. Eng. **27**(3), 561–570 (2002).

[19] P. D. Mourad and D. R. Jackson, "High frequency sonar equation models for bottom backscatter and forward loss," in *Proceedings Oceans '89 (Marine Technology Society and IEEE)*, 1989, pp. 1168–1175.

[20] M. A. Ainslie, "Reflection and transmission coefficients for a layered fluid sediment overlying a uniform solid substrate," J. Acoust. Soc. Am. **99**, 893–902 (1996).

[21] D. Tang, "Fine-scale measurements of sediment roughness and subbottom variability," IEEE J. Ocean. Eng. **29**, 929–939 (2004).

[22] J. H. Miller, L. R. Bartek, G. R. Potty, D. Tang, J. Na, and Y. Qi, "Sediments in the East China Sea," IEEE J. Ocean. Eng. **29**(4), 940–951 (2004).

[23] P. A. van Walree, J. Tęgowski, C. Laban, and D. G. Simons, "Acoustic seafloor discrimination with echo shape parameters: A comparison with the ground truth," Cont. Shelf Res. **25**, 2273–2293 (2005).

[24] T. Jenserud and D. Ottesen, "Analysis of bottom samples from Vestfjorden

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

van Walree et al.: Mean grain size mapping   2565

collected during RUMBLE first sea trial," FFI/RAPPORT-2002/05018, December 2002.

[25] T. Jenserud and D. Ottesen, "Analysis of bottom samples collected from the Norwegian Trench during RUMBLE second sea trial," FFI/RAPPORT-2003/01104, January 2004.

[26] M. Kimura and T. Tsurumi, "Characteristics of acoustic wave reflection from the transition layer of surficial marine sediment," in *IEEE Underwater Technology 2002*, Tokyo, pp. 225–230.

[27] M. J. N. Bergman and M. Hup, "Direct effects of beam trawling on macrofauna in a sandy sediment in the southern North Sea," ICES J. Mar. Sci. **49**, 5–11 (1992).

[28] A. Linnane, B. Ball, B. Munday, B. van Marlen, M. Bergman, and R. Fonteyne, "A review of potential techniques to reduce the environmental impact of demersal trawls," Irish Fisheries Investigations **7**, 1–39 (2000).

[29] P. Schwinghamer, D. C. Gordon Jr., T. W. Rowell, J. Prena, D. L. McKeown, G. Sonnichsen, and J. Y. Guigné, "Effects of experimental otter trawling on surficial sediment properties of a sandy bottom ecosystem on the Grand Banks of Newfoundland," Conserv. Biol. **12**(6), 1215–1222 (1998).

[30] I. Tuck, S. Hall, M. Robertson, E. Armstrong, and D. Basford, "Effects of physical trawling disturbance in a previously unfished sheltered Scottish sea loch," Mar. Ecol.: Prog. Ser. **162**, 227–242 (1998).

[31] M. J. Kaiser and B. E. Spencer, "The effects of beam trawl disturbance on infaunal communities in different habitats," J. Anim. Ecol. **65**, 348–358 (1996).

[32] F. Pranovi, S. Raicevich, G. Franceschini, M. G. Farrace, and O. Giovanardi, "Rapido trawling in the northern Adriatic Sea: Effects on benthic communities in an experimental area," ICES Journal of Marine Science **57**(3), 517–524 (2000).

[33] E. J. Brown, B. Finney, M. Dommisse, and S. Hills, "Effects of commercial otter trawling on the physical environment of the southeasthern Bering Sea," Cont. Shelf Res. **25**, 1281–1301 (2005).

A stochastic representation of environmental uncertainty and its coupling to acoustic wave propagation in ocean waveguides

Steven Finette
Acoustics Division, Naval Research Laboratory, Washington, D.C. 20375

It is argued that a quantitative measure of incomplete environmental knowledge or information (i.e., environmental uncertainty) should be included in any simulation-based predictions linked to acoustic wave propagation. A method is then proposed to incorporate environmental uncertainty directly into the computation of acoustic wave propagation in ocean waveguides. In this regard, polynomial chaos expansions are chosen to represent uncertainty in both the environment and acoustic field. The sound-speed distribution and acoustic field are therefore generalized to stochastic processes, where uncertainty in the field is interpreted in terms of its statistical moments. Starting from the narrow angle parabolic approximation, a set of coupled differential equations is derived in which the coupling term links incomplete environmental information to the corresponding uncertainty in the acoustic field. Propagation of both the field and its uncertainty in an isospeed waveguide is considered as an example, where the sound speed is described by a random variable. The first two moments of the field are computed explicitly and compared to those obtained from independent Monte Carlo solution of the conventional (deterministic) parabolic equation that describes the acoustic wave properties. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2335425]

I. INTRODUCTION

Numerical simulation of acoustic wave propagation in ocean waveguides has served as an indispensable tool for the exploration, interpretation, and prediction of acoustic field properties in complex ocean environments. An application of simulation techniques to predict field properties or the subsequent effect of these properties on sonar system performance represents a particular example of simulation-based prediction. The validity of simulation-based prediction depends, to a large extent, on the assumption that either all environmental information necessary for the solution of the problem is known or, if this information is only partially available, that the resulting uncertainty in one's knowledge of the environment can be objectively quantified and included in the result.[1] If neither of these conditions is satisfied, the prediction is flawed and conclusions or decisions based on the results are of questionable validity. Since environmental knowledge is always incomplete for acoustic propagation under realistic ocean waveguide conditions, it has recently been argued that a prediction of acoustic field properties would benefit from the inclusion of a quantitative measure of environmental uncertainty.[2,3] The idea of embedding uncertainty directly into the simulation framework and elevating its status to a subject worth studying on its own merits represents a paradigm shift in a wide range of disciplines.[4–6] The purpose of this paper is to describe a rather general approach for the direct inclusion of environmental uncertainty into the dynamical formulation of acoustic propagation, with the goal of improving simulation-based prediction of the acoustic field. An example of the method is given for a simple stochastic waveguide environment in or-

der to illustrate the approach in a setting where numerical implementation issues are not a significant factor. Before discussing a method for the objective assessment of environmental uncertainty and its subsequent effect on the acoustic field, the term "uncertainty" and the phrase "environmental knowledge" need some elaboration.

A general, rigorous definition of uncertainty is rather difficult to present even though one has an intuitive sense of the term's meaning in the above comments. An informal and somewhat more restrictive definition is relevant for this discussion: uncertainty, in the context of simulation-based prediction, represents the lack of complete knowledge (information) concerning a property or set of properties that must be specified in order to model and simulate a physical system. The system addressed here is comprised of an acoustic field and the waveguide environment that supports the propagating field. In a simulation of acoustic wave propagation, the representation of, say, both the sound-speed field and bathymetry by correlation functions or power spectra is an acknowledgment that complete information about the system response is not known. The spectra or correlation functions then represent the current state of available knowledge about the environment and can be used to place bounds on the set of possible system responses that can be computed. One can interpret the spectrum as a reflection of uncertainty in the environmental description of the system, with the implication that a range of possible system responses should be simulated in order to fully characterize the propagation consistent with the available information. While the range of possible simulated responses due to incomplete environmental information might be considered as a measure of natural

variability, the source of this variability is an information deficit rather than some physical mechanism governing system behavior. Though there are features in common, one should not consider natural variability in measured data as an exact counterpart to uncertainty in simulation-based prediction. A consequence of this uncertainty is to effectively introduce additional degrees of freedom into the simulation, and their effect on propagation should be taken into account in simulation-based prediction schemes. In the following sections a general method is considered for explicitly including these additional degrees of freedom within the dynamics. The emphasis is on the influence of environmental uncertainty in the specification of the sound-speed distribution and its subsequent effect on acoustic field propagation. The emphasis on sound-speed uncertainty covers a broad range of practical scenarios commonly encountered in ocean acoustics simulation. A sound-speed distribution can be represented by the set of numerical values of the field on a discrete computational grid over a time interval $[t_{\text{initial}}, t_{\text{final}}]$, during which the system dynamics is evolved. A simulated acoustic field propagating through this distribution might accumulate significant natural variability in amplitude and phase. Suppose, however, that the sound-speed information is only partially specified at a subset of grid points and then interpolated to other points on the grid or, instead, it is specified through second-order statistical information (correlation function or power spectrum) where higher order information is neglected (but not zero). In these cases environmental uncertainty occurs in the specification of the sound-speed distribution and it should be reflected in a range of possible simulated acoustic field responses, in addition to any natural variability that can occur as the field propagates through the waveguide.

It is important to note that uncertainty, as considered above, is quite distinct from numerical errors that inevitably arise when a mathematical formulation of a propagation model is discretized, implemented on a computer, and solved with finite precision arithmetic. Errors arising from this processing have been studied in numerical analysis for over 50 years and constitute a rather mature subject area. Here it is emphasized that, in addition to numerical errors, incomplete knowledge of the waveguide's environment may significantly limit the ability to make valid predictions from the results of numerical simulation unless the effects of environmental uncertainty are included in the simulation process.

A probabilistic framework is described in the following sections to quantify incomplete environmental knowledge and the resulting uncertainty in the acoustic field. Probability distributions are interpreted here as natural representations of uncertainty. In this sense the system—comprised of both the environment and the dynamics—loses its deterministic structure and is described probabilistically. Statistical moments are considered appropriate quantitative metrics for describing uncertainty in terms of the spread of a distribution. While moments are not a unique indicator of uncertainty, they are chosen here because the first few moments of the acoustic field have simple physical interpretations. It is therefore natural to extend the notion of environmental parameters and fields, treated deterministically in a single numerical simula-

tion, and replace them with random variables and stochastic processes that naturally include uncertainty. Note that this generalization does not necessarily imply that the sound speed and acoustic field are true stochastic processes but rather, for the purpose of including uncertainty in the dynamics, it is convenient to treat them mathematically as if they were stochastic processes. In this manner, a probability-weighted representation of possible sound-speed distributions is considered as well as the subsequent effect of these distributions on the acoustic field (also probability weighted). While the distinction between uncertainty and variability becomes somewhat blurred in this view, it does not affect the quantitative analysis presented here.

The approach represents these random variables and processes explicitly in terms of stochastic basis expansions. Doing so implies that the deterministic wave equation is replaced by a stochastic wave equation. The solution of this equation yields the stochastic process representing the acoustic field; statistical moments of the field can be computed directly from the solution. Additional complexity is involved in solving a stochastic differential equation, but it is mitigated by several factors that may yield potential advantages over alternative approaches for obtaining moments of the acoustic field, such as Monte Carlo sampling[7] and moment equation methods.[8,9] Below, a brief comparison is made between these well-known methods and the formulation discussed in the following sections.

Monte Carlo techniques are commonly applied to a deterministic wave equation whose parameters and/or fields are specified by repeated sampling from probability distributions on these quantities in order to obtain members of an environmental ensemble. Using members of the ensemble, multiple deterministic realizations of the acoustic field are computed from the equation and then postprocessed to obtain the moments. The basic methodology is relatively straightforward and can be applied to any wave equation describing the acoustic field. One problem with this approach is that large numbers of realizations are required and the rate of convergence is slow because it follows the law of large numbers with standard deviation proportional to $1/\sqrt{R}$, where R is the number of realizations.[7,10] In addition, the convergence is not monotone. For relatively simple wave equations and environments, the CPU time necessary to generate the set of realizations is manageable. However, for complex problems such as the case where the sound-speed distribution is determined directly from fluid dynamics computations, the CPU time can exceed the capabilities of modern supercomputers.[10] The individual realizations give little direct insight into the field statistics and must be further processed to obtain the moments. The stochastic expansion approach using polynomial chaos (PC) basis functionals, summarized in Sec. II, effectively includes all realizations of the stochastic pressure field without resorting to explicit sampling from probability distributions and the subsequent computation of multiple solutions of the wave equation. This method often yields comparable results that are several orders of magnitude faster than those obtained from Monte Carlo sampling.[11,12] Individual realizations can be obtained directly from the random process computed using the PC

expansion approach. In contrast to a realization-based sampling method, the compact formulation involving PC expansions can be very useful for the analysis of multiple subsystems that are coupled together, with uncertainty in one system transferred to another. An example is a three-component model. In the first subsystem, environmental uncertainty in the temperature and salinity distributions is introduced into the fluid dynamic equations of motion and a phenomenological equation relating sound speed to temperature and salinity. Uncertainty linked to the sound-speed distribution is then transferred to the second subsystem consisting of acoustic propagation through the uncertain environment. The third subsystem is comprised of signal processing applied to the resulting stochastic acoustic field received on a sensor array.

Introduced into ocean acoustics by Dozier and Tappert,[13,14] moment equation formulations of wave propagation develop separate evolution equations governing each moment of the acoustic field rather than the field itself. A number of different equations have been proposed for propagating full field or modal moments, with each approach dependent on several mathematical and physical assumptions having varying degrees of generality and rigor. These efforts have included, for example, both deep[13,15] and shallow[16,17,20] water environments with formulations in terms of either full field[18,19] or modal[20–22] moments. The complexity of these equations typically increases with increasing moment order. In contrast, polynomial chaos expansions compute the moments from the random process itself; the process is obtained explicitly as the solution to a set of coupled differential equations that include both the propagation dynamics and coupling between environmental and acoustic field uncertainties. The moments are not directly propagated in the PC formulation, but are obtained indirectly through this solution, and separate equations for the individual moments are not necessary. From the point of view of simulation-based prediction, the moments derived from the polynomial chaos method are interpreted as measures of the effect of environmental uncertainty on acoustic field structure, rather than as measures of natural variability associated with the field. The compact form of the solution allows for some flexibility in both additional manipulation and further analysis; as mentioned above, such a manageable representation is quite useful when analyzing a system in which uncertainty is transferred between subsystems.

Polynomial chaos expansions are obtained from a theorem in the theory of random processes under the major assumption that the process in question is second order, i.e., has finite second moment. This assumption represents a very reasonable property satisfied by a large class of physical systems. Truncation of the resulting infinite series expansions is necessary for numerical computation, and ensures a form of closure.[23] Information on the higher order moments is retained in the truncated expansions. While these series expansions provide the best approximation of the process (in the sense of minimizing the mean-squared error) given the particular level of approximation, the implications of truncation on the numerical accuracy and convergence rate for the moments depends on the type of problem, the random process to

be approximated, and the choice of basis functional; these issues are a subject of current research.[23–25] The Markov approximation is not explicitly assumed in the polynomial chaos formulation, though the correlation structure of the medium plays an important role in determining the rate of convergence.[23] The polynomial chaos method can be considered as a mathematical formalism for representing uncertainty though it does not introduce new physics into the problem. Instead, all the physics and the coupling between environmental uncertainty and the resulting acoustic field uncertainty are present in a generalized "wave equation" for the expansion coefficients that results from applying this formalism.

Section II discusses the general theoretical framework for polynomial chaos expansions. These expansions are applied in Sec. III to develop equations that describe propagation of the acoustic field in the presence of environmental uncertainty, when propagation is approximated by the narrow-angle parabolic equation. The equations are then solved for the case of an isospeed stochastic waveguide, and the results are compared to independent Monte Carlo estimates of the first and second field moments obtained by computing multiple field realizations from the deterministic parabolic equation. Section IV presents a summary and conclusions.

II. REPRESENTATION OF UNCERTAINTY IN TERMS OF STOCHASTIC BASIS EXPANSIONS

This section describes a mathematical framework for representing incomplete environmental knowledge and linking it to uncertainty in the acoustic field. While the discussion is not rigorous, enough details are presented so that the generality of the method is clear and references to more rigorous presentations are included. Before considering stochastic expansions, it is helpful to note that the representation of random processes by stochastic basis expansions may be considered as a generalization of the way one can represent deterministic functions by polynomial expansions. The Stone-Weierstrass approximation theorem[26] allows for the expansion a square-integrable function $f(x)$ in terms of a complete set of basis functions (in this case polynomials in the variable x). A monomial expansion is simply given by $f(x) = \sum_{m=0}^{\infty} a_m x^m$. The monomials form a complete set of functions in a Hilbert space. While these polynomials are not orthonormal, they are linearly independent. The basis functions can be made orthonormal by application of the Gram-Schmidt procedure. For polynomials of more than one variable, a deterministic multinomial expansion can be written as a multiple sum with the general form $f(x_1, x_2, \cdots, x_n) = \sum_{m_1 \ldots m_n}^{\infty} a_{m_1 m_2 \ldots m_n} x_1^{m_1} x_2^{m_2} \cdots x_n^{m_n}$.

As mentioned in the Introduction, uncertainty is described here within the framework of probability theory. The approach assumes that uncertainty in the acoustic field is attributed to incomplete environmental knowledge of the sound-speed distribution, and can be quantified using the statistical moments of the field. Therefore, environmental parameters and fields are not described by numbers or deterministic functions but rather in terms of random variables or

stochastic processes. This description implies that the pressure field is also specified by a stochastic process whose moments carry information about the effect of environmental uncertainty. A natural probabilistic method for describing uncertainty has evolved from research in stochastic finite elements[27] that builds on earlier mathematical work by Wiener, who dealt with the problem of treating Brownian motion as a stochastic process.[28] His work introduced a particular type of stochastic basis expansion, termed polynomial chaos expansions, to describe random processes. It is worth mentioning that the word "chaos" was invoked by Wiener before its more familiar use in nonlinear dynamics; the two subjects, however, are independent of each other. The research in stochastic finite elements significantly extended Wiener's original work by systematically applying polynomial chaos expansions in a dynamical framework,[27] merging these expansions with stochastic differential equations[10–12] and interpreting them in terms of uncertainty in the system response.[29–31]

Assume that the environmental parameters or fields are describable by second-order random variables or stochastic processes, i.e., they have finite variance. For generality, it is convenient to define a probability space[27,32] as a triple (Π, Θ, Ξ), where Π denotes a sample space (the set of all elementary distinguishable outcomes of a random trial or experiment), Θ is an event space (or sigma-algebra) consisting of a suitably restricted set of subsets of Π, and Ξ is a probability (or probability measure) defined on all the subsets in Θ. Let $\theta \in \Theta$, i.e., θ represents a possible outcome (random event, realization) of an experiment associated with Θ. The probability measure is a set function operating on Θ that assigns a probability to each event $\theta = \theta_\alpha$. In the context of ocean acoustics, Π could represent the set of possible complex pressure field values that can be measured in continental shelf environments throughout the year. An event space Θ might then contain subsets of possible field values that can be measured during the summer months off the coasts of New Jersey and Oregon for a particular frequency band and source level. The quantity $\theta = \theta_\alpha$ then represents a particular pressure distribution (realization) measured under the above conditions, with a weighting or likelihood of occurrence determined by the probability measure Ξ.

Let $\mu(\theta)$ be a second-order process considered as a function of the arbitrary event θ. A theorem in the theory of stochastic processes[33] states that one can always represent this process as a series expansion, $S(\theta)$, in terms of a complete set of orthogonal random polynomials $\{\Phi_d, d = 1, 2, \dots\}$ that are functions of random variables $\xi_{j_i}(\theta)$ for positive integers $j_1, j_2, \dots$. It will be apparent in what follows that including the quantity θ is a useful notation for distinguishing between stochastic and deterministic quantities. The expansion is guaranteed to converge to $\mu(\theta)$ in the mean-squared sense,

$$\langle (\mu(\theta) - S_N(\theta))^2 \rangle \to 0 \quad \text{as } N \to \infty, \tag{1}$$

where S_N is the Nth partial sum of the series S and $\langle \rangle$ represents an ensemble average with respect to a probability measure to be discussed shortly. Explicitly, the expansion for the

second-order space-time process μ can be written in the form[27]

$$\mu(\mathbf{r}, t; \theta) = b_o(\mathbf{r}, t)\Phi_o + \sum_{j_1=1}^{\infty} b_{j_1}(\mathbf{r}, t)\Phi_1(\xi_{j_1}(\theta))$$

$$+ \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{j_1} b_{j_1 j_2}(\mathbf{r}, t)\Phi_2(\xi_{j_1}(\theta), \xi_{j_2}(\theta))$$

$$+ \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{j_1} \sum_{j_3=1}^{j_2} b_{j_1 j_2 j_3}(\mathbf{r}, t)\Phi_3(\xi_{j_1}(\theta), \xi_{j_2}(\theta), \xi_{j_3}(\theta)) + \cdots.$$

$$\tag{2}$$

The functionals Φ_d are named the polynomial chaos of order d, and specify a complete orthogonal basis in a Hilbert space of second-order random variables.[27] The polynomial chaos are multivariate orthogonal functionals since they are functions of the random variables ξ_{j_i} describing the uncertainty. The sum in Eq. (2) is carried out according to the order d in such a manner that lower order terms are evaluated first. The word "order" refers to the highest power of the random polynomials appearing in the basis functionals; it is not to be confused with the use of the same word in the phrase "second-order process." Equation (2) describes a second-order stochastic process as an infinite series expansion in terms of sums of products of stochastic basis functionals and deterministic expansion coefficients $b_{j_1 j_2}, \dots$. Such a separation between stochastic and deterministic quantities is a special case of a general result in the theory of second-order stochastic processes, in which a spectral representation can be found that separates the process into stochastic and deterministic contributions.[27] The series in Eq. (2) has been written as a space-time process in which the coefficients contain the space, $\mathbf{r} = (x, y, z)$, and time, t, dependence and the basis functionals Φ_d are independent of these variables. For the example discussed here, the random variables $\xi_{j_1}, \xi_{j_2}, \dots$ are linked to uncertainty in the sound-speed distribution. In practical applications, the infinite series representation in Eq. (2) must be truncated to a finite number of terms for any order. In general, if the highest order polynomial is d, then the total number of expansion terms for an n-dimensional random vector $\boldsymbol{\xi}$ will contain $(n+d)!/n!d!$ terms.[11,12]

The basis functionals can be explicitly identified from the Wiener-Askey class of multivariate orthogonal functionals;[12] these include (but are not limited to) the familiar Hermite, Laguerre, and Legendre polynomials. Once the deterministic expansion coefficients, $b_{j_1 j_2 \dots}$, are computed, Eq. (2) contains a complete probabilistic description of the stochastic process and associated uncertainty. The set of stochastic functionals $\{\Phi_d\}$ forms a complete orthogonal basis set in this Hilbert space in the sense that the ensemble average of a pair of basis functionals is given by[12,27]

$$\langle \Phi_d \Phi_{d'} \rangle \equiv \int_\Theta \Phi_d(\boldsymbol{\xi})\Phi_{d'}(\boldsymbol{\xi})W(\boldsymbol{\xi})d\boldsymbol{\xi} = \langle \Phi_d^2 \rangle \delta_{dd'}, \tag{3}$$

where $W(\boldsymbol{\xi})$ is a probability density function defined through Ξ, and the integral is over the support of the random vector $\boldsymbol{\xi} = \{\xi_{j_1}, \xi_{j_2} \dots\}$. The choice of density function or probability

measure depends on the particular class of polynomial chaoses[11,12,27] chosen for the problem. The multivariate density function acts as a weighting factor on the random variables $\boldsymbol{\xi}$, that characterize uncertainty in the system. For example, if $\boldsymbol{\xi}$, represents a multidimensional Gaussian random variable, then the corresponding polynomial chaos basis functionals are the Hermite polynomials and the weighting (density) function W is a multidimensional Gaussian density function. On the other hand, Laguerre and Legendre polynomials are weighted by gamma and uniform distributions, respectively. From the point of view of uncertainty and its propagation through the system, the basis functionals link the relative weightings of both the environmental and acoustic field realizations that comprise the process, while the deterministic coefficients limit the set of possible realizations through the constraints imposed by the dynamics of wave propagation. These points will be clarified in the next section in terms of an example. Monte Carlo simulation represents a special case of Eq. (2) where a particular realization or outcome of a trial $\theta = \theta_\alpha$ is selected.[11] In effect, polynomial chaos expansions allow for all possible realizations to be computed simultaneously and described in a relatively compact form. An equivalent determination of individual realizations using Monte Carlo sampling can be accomplished for each θ_α directly from Eq. (2), provided the coefficients are known.

For notational convenience the series in Eq. (2) is written in the simplified notation[12,27] $\mu(\theta) = \Sigma_{q=0}^{\infty} c_q \Lambda_q(\boldsymbol{\xi}(\theta))$ where a one-to-one mapping relates the terms in this expression with those of Eq. (2). This form will be used throughout the text and involves a renumbering of Eq. (2) with the polynomials of lower order counted first; except for the indexing convention, Λ and Φ are identical,[12,27] though note that q does not generally represent the order of the expansion. For example, a polynomial chaos expansion of order 3 and random dimension 2 contains two random variables and a total of ten terms. It can be written compactly in the form $\mu(\boldsymbol{r}, t; \theta) = \Sigma_{q=0}^{9} c_q(\boldsymbol{r}, t) \Lambda_q(\boldsymbol{\xi}(\theta))$, where the identification of terms in this sum with those of Eq. (2) is given by e.g., $c_3 \Lambda_3 = b_2 \Phi_1(\xi_2), c_4 \Lambda_4 = b_{11} \Phi_2(\xi_1, \xi_1), c_7 \Lambda_7 = b_{211} \Phi_3(\xi_2, \xi_1, \xi_1)$, etc. The bases Λ_q also satisfy the orthogonality relation given by Eq. (3). An explicit example of the use of this polynomial chaos expansion is given in Sec. III after dynamical equations are developed to link both environmental and acoustic field uncertainty with propagation.

III. UNCERTAINTY EQUATIONS IN THE NARROW-ANGLE PARABOLIC APPROXIMATION

A. General formulation for arbitrary sound-speed distribution

The mathematical framework discussed in the previous section is applied here to develop equations for propagation within an uncertain waveguide environment, where the acoustic field is described by a narrow-angle parabolic approximation to the Helmholtz equation. In this approximation, the (deterministic) parabolic equation has the form[34]

$$2ik_o \frac{\partial \psi(r,z)}{\partial r} + \frac{\partial^2 \psi(r,z)}{\partial z^2} + k_o^2(N^2(r,z) - 1)\psi(r,z) = 0, \quad (4)$$

where the acoustic pressure $P(r,z)$ in range r and depth z is related to an envelope function $\psi(r,z)$ by the relation $P(r,z) \approx \sqrt{(2/\pi k_o r)}e^{i(k_o r - (\pi/4))}\psi(r,z)$ in the far-field approximation. The function $N(r,z)$ is the index of refraction, $c_o/c(r,z)$, relative to a reference wave number $k_o = \omega/c_o$ and $i = \sqrt{-1}$. To account for environmental uncertainty in the water column, the deterministic sound-speed field is now generalized to a stochastic process by the ansatz: $c(r,z;\theta) \Rightarrow \bar{c}(r,z) + \delta c(r,z;\theta)$. The function $\bar{c}(r,z)$ describes the deterministic mean field; $\delta c(r,z;\theta)$ represents the stochastic contribution with θ denoting an arbitrary element of Θ containing possible realizations of the sound-speed field weighted by the probability measure Ξ. For ocean acoustics problems the sound-speed perturbations represented by δc are quite small relative to $\bar{c}$, and it is assumed here that $\langle (\frac{\delta c}{\bar{c}})^2 \rangle \ll 1$, where $\langle \rangle$ denotes ensemble average.

Substitution of $c(r,z;\theta)$ into the index of refraction, and keeping only the first-order term in a binomial expansion of the sound speed, gives

$$n^2(r,z;\theta) \approx c_o^2 \left[\frac{1}{\bar{c}^2(r,z)} \left(1 - \frac{2\delta c(r,z;\theta)}{\bar{c}(r,z)} \right) \right].$$

Therefore, Eq. (4) is generalized to the following stochastic differential equation:

$$2ik_o \frac{\partial \psi(r,z;\theta)}{\partial r} + \frac{\partial^2 \psi(r,z;\theta)}{\partial z^2} + k_o^2 \left(\frac{c_o^2}{\bar{c}^2(r,z)} - 1 \right) \psi(r,z;\theta)$$
$$- \frac{2c_o^2 k_o^2 \delta c(r,z;\theta)}{\bar{c}^3(r,z)} \psi(r,z;\theta) = 0. \quad (5)$$

Note that $\psi(r,z) \Rightarrow \psi(r,z;\theta)$, because ψ is now a function of the stochastic process $\delta c(r,z;\theta)$. Assuming both the sound-speed perturbation and the envelope function are second-order stochastic processes, each process can be expanded in terms of polynomial chaos basis functionals $\Lambda(\boldsymbol{\xi}(\theta))$,

$$\delta c(r,z;\theta) = \sum_{s=0}^{\infty} \alpha_s(r,z) \Lambda_s(\boldsymbol{\xi}(\theta)), \quad (6a)$$

$$\psi(r,z;\theta) = \sum_{q=0}^{\infty} \gamma_q(r,z) \Lambda_q(\boldsymbol{\xi}(\theta)), \quad (6b)$$

where $\alpha_s(r,z), \gamma_q(r,z)$ are the deterministic uncertainty coefficients for the sound speed and envelope processes, respectively. Substituting Eqs. (6) into Eq. (5), multiplying the result by $\Lambda_l(\boldsymbol{\xi}(\theta))$, and projecting onto this basis by ensemble averaging each term yields

$$2ik_o\sum_{q=0}^{\infty}\frac{\partial\gamma_q(r,z)}{\partial r}\langle\Lambda_q\Lambda_l\rangle + \sum_{q=0}^{\infty}\frac{\partial^2\gamma_q(r,z)}{\partial z^2}\langle\Lambda_q\Lambda_l\rangle$$

$$+ k_o^2\left[\left(\frac{c_o^2}{\bar{c}^2(r,z)}-1\right)\right]\sum_{q=0}^{\infty}\gamma_q(r,z)\langle\Lambda_q\Lambda_l\rangle$$

$$-\frac{2c_o^2k_o^2}{\bar{c}^3(r,z)}\sum_{q=0}^{\infty}\sum_{s=0}^{\infty}\alpha_s(r,z)\gamma_q(r,z)\langle\Lambda_q\Lambda_s\Lambda_l\rangle = 0. \quad (7)$$

It is assumed for the forward problem that the stochastic process describing the uncertainty in the sound-speed field is known or can be estimated from data (i.e., the coefficients α_s are known). By applying the orthogonality condition on the chaos basis functionals [Eq. (3)] to the terms in Eq. (7), one then obtains an infinite set of deterministic, coupled partial differential equations for the uncertainty coefficients of the envelop field,[3]

$$2ik_o\frac{\partial\gamma_l(r,z)}{\partial r} + \frac{\partial^2\gamma_l(r,z)}{\partial z^2} + k_o^2\left[\frac{c_o^2}{\bar{c}^2(r,z)}-1\right]\gamma_l(r,z)$$

$$-\frac{2c_o^2k_o^2}{\bar{c}^3(r,z)\langle\Lambda_l^2\rangle}\sum_{q=0}^{\infty}\sum_{s=0}^{\infty}\alpha_s(r,z)\gamma_q(r,z)\langle\Lambda_q\Lambda_s\Lambda_l\rangle = 0.$$

$$(8)$$

The last term in Eqs. (8) represents the coupling of uncertainty between the sound-speed distribution and the envelope function determining the acoustic field. These coupled equations are rather general, describing the propagation of the envelope and its associated uncertainty constrained by both the dynamics imposed by the narrow-angle parabolic equation and the stochastic sound-speed environment. The particular propagation characteristics included in the formulation (e.g., scattering, dispersion, and attenuation) depend on the choice of wave equation and boundary conditions. Solving Eqs. (8) for the uncertainty coefficients $\gamma(r,z)$ $\equiv[\gamma_1,\gamma_2,\ldots,\gamma_\infty]$ completely determines the process describing the stochastic pressure field and its corresponding uncertainty, since

$$P(r,z;\theta) \approx \sqrt{\frac{2}{\pi k_o r}}e^{i(k_o r-(\pi/4))}\psi(r,z;\theta)$$

$$= \sqrt{\frac{2}{\pi k_o r}}e^{i(k_o r-(\pi/4))}\sum_{q=0}^{\infty}\gamma_q(r,z)\Lambda_q(\boldsymbol{\xi}(\theta)). \quad (9)$$

Note that the dummy index l in Eq. (8) has been replaced here by the dummy summation index q in Eq. (9).

B. Solution for the stochastic acoustic field in an isospeed waveguide

An analytic solution of Eq. (8) together with several moments of the acoustic field are computed here for an idealized stochastic waveguide in order to illustrate the theory. An isospeed environment with a pressure release surface and a rigid bottom is considered, in which the sound-speed field is chosen to be a Gaussian random variable. This environment is shown in Fig. 1 for a waveguide of water depth D with a source located at a depth z_0 below the surface. A
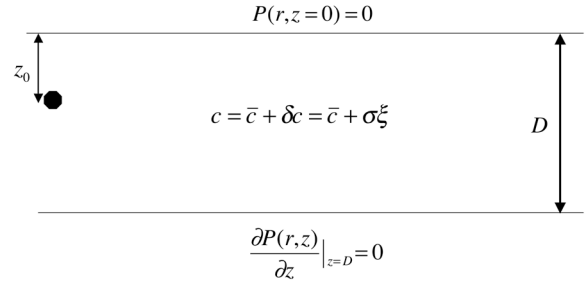
FIG. 1. Spatially homogeneous stochastic waveguide environment used in the analysis. The boundary conditions on the surface and bottom are pressure release and rigid bottom, respectively.

stochastic waveguide is equivalent to a family of deterministic waveguide environments[35] where each member corresponds to a single realization of the sound-speed field and the likelihood of any realization is determined by a probability distribution. In this particular case each member is drawn from a Gaussian distribution. While the model does not represent a realistic waveguide environment, its simplicity allows us to compute the moments explicitly and illustrate the methodology in a transparent manner without the additional complexity of dealing with numerical issues related to Eqs. (8). Numerical solutions of Eqs. (8) for spatially varying sound-speed environments are beyond the scope of this paper.

The Gauss-Hermite functionals will be used as the specific bases in the polynomial chaos expansions to describe both the stochastic acoustic field and sound-speed distribution. This is not an essential choice and different basis sets could be used, for example, to separately specify the environment and acoustic processes. While convergence of the series expansion in Eq. (2) is assured for second-order processes,[12,33] the rate of convergence is dependent on the choice of basis for a given environment. For a single random variable ξ the Gauss-Hermite polynomials, $\Lambda_n(\boldsymbol{\xi})\rightarrow H_n(\xi)$, satisfy the orthogonality condition[36]

$$\int_{-\infty}^{\infty}e^{-\xi^2}H_n(\xi)H_{n'}(\xi)d\xi = \pi^{1/2}2^n n!\,\delta_{nn'}, \quad (10)$$

for integer n and with $W(\boldsymbol{\xi})=e^{-\xi^2}$. Isospeed sound-speed perturbations δc are represented here as a zero-mean, Gaussian random variable with unit variance, $\xi\sim\mathbf{N}(0,1)$. The sound-speed distribution can then be expressed by $c(r,z)\Rightarrow c=\bar{c}+\sigma\xi$, where σ is the standard deviation of the sound-speed fluctuations. This implies that the PC series representation [Eq. (2)] reduces to a one-dimensional expansion of order d. Note that this dimensionality does not refer to the number of space dimensions in the problem; here, a two-dimensional waveguide is considered so that the expansion coefficients are functions of both range and depth. The Gauss-Hermite polynomials need to be scaled since the weighting function appearing in Eq. (10) does not represent a unit variance distribution. Replacing ξ by $\xi/\sqrt{2}$, the scaled basis functions, denoted by $He_n(\xi)$, can be written in the form $He_n(\xi)=(2^{-n/2}/\sqrt{n!})H_n(\xi/\sqrt{2})$. Using Eq. (10), this choice of PC basis functional satisfies the orthonormality condition

Steven I. Finette: Stochastic representation of uncertainty

$$\langle He_n(\xi)He_{n'}(\xi)\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\xi^2/2}He_n(\xi)He_{n'}(\xi)d\xi = \delta_{nn'},$$

$$(11)$$

consistent with the random sound-speed perturbation having zero mean and unit variance. The condition expressed in Eq. (11) is a particular case of Eq. (3) for which the basis functionals have been normalized to unity. These basis functionals can be determined explicitly from the recursion relation

$$\xi He_n(\xi) = \sqrt{n+1}He_{n+1}(\xi) + \sqrt{n}He_{n-1}(\xi),$$

$$(12)$$

subject to the constraints $He_{-1}=0, He_0=1$. This expression is obtained using the recursion relation[26,36] for $H_n(\xi)$ and its relationship with $He_n(\xi)$ given above. The first few orthonormal Gauss-Hermite functionals for one-dimensional polynomial chaos of order $n=1,2,3$ are then given explicitly by

$$He_1(\xi) = \xi, \quad He_2(\xi) = \frac{\xi^2-1}{\sqrt{2}}, \quad He_3(\xi) = \frac{\xi^3-3\xi}{\sqrt{6}}. \quad (13)$$

By substituting the Gauss-Hermite bases into Eq. (6a), the sound-speed perturbation term can be written as $\delta c = \alpha_o He_o + \alpha_1 He_1 + \cdots = \sigma\xi$. Only the second term survives with the identification of α_1 as the standard deviation, σ, of the perturbation consistent with the $N(0,1)$ assumption. Substituting this nonzero term into Eqs. (8) eliminates the sum over s and gives the following set of coupled equations for the lth uncertainty coefficient:

$$2ik_o\frac{\partial\gamma_l(r,z)}{\partial r} + \frac{\partial^2\gamma_l(r,z)}{\partial z^2} + k_o^2\left[\frac{c_o^2}{\bar{c}^2}-1\right]\gamma_l(r,z)$$

$$-\frac{2c_o^2k_o^2\sigma}{\bar{c}^3\langle He_l^2\rangle}\sum_{q=0}^{\infty}\gamma_q(r,z)\langle He_q\xi He_l\rangle = 0. \quad (14)$$

Using Eq. (11) and Eq. (12), the ensemble averages appearing in the coupling term can be explicitly computed, with the result that any γ_l is determined by the partial differential recursion equation

$$2ik_o\frac{\partial\gamma_l(r,z)}{\partial r} + \frac{\partial^2\gamma_l(r,z)}{\partial z^2} + D\gamma_l(r,z) - E\sqrt{l}\gamma_{l-1}(r,z)$$

$$-E\sqrt{l+1}\,\gamma_{l+1}(r,z) = 0, \quad (15)$$

where $D\equiv k_o^2((c_o^2/\bar{c}^2)-1)$ and $E\equiv 2c_o^2k_o^2\sigma/\bar{c}^3$. It is convenient to write this equation in vector-matrix form for $\boldsymbol{\gamma}$,

$$2ik_o\frac{\partial\boldsymbol{\gamma}}{\partial r} + \frac{\partial^2\boldsymbol{\gamma}}{\partial z^2} + [A]\boldsymbol{\gamma} = 0, \quad (16)$$

where the uncertainty coupling matrix $[A]$ is symmetric and tridiagonal. For a sixth-order expansion the matrix is given explicitly by

$$[A] = \begin{bmatrix} D & -E & 0 & 0 & 0 & 0 \\ -E & D & -\sqrt{2}E & 0 & 0 & 0 \\ 0 & -\sqrt{2}E & D & -\sqrt{3}E & 0 & 0 \\ 0 & 0 & -\sqrt{3}E & D & -2E & 0 \\ 0 & 0 & 0 & -2E & D & -\sqrt{5}E \\ 0 & 0 & 0 & 0 & -\sqrt{5}E & D \end{bmatrix}.$$

$$(17)$$

The off-diagonal components represent the "magnitude" of the environmental uncertainty since these matrix elements are functions of σ. In general, $[A]$ is a function of spatial coordinates associated with the sound-speed field $c(r,z)$. This is clear from Eqs. (8) and prevents an analytic solution in cases of arbitrary sound speed. However, Eq. (16) has constant coefficients by virtue of the assumed independence of $\bar{c}+\delta c$ on position and it can be solved exactly by performing a similarity transformation to diagonalize the coupling matrix.[37]

Define $[G]=[\{g_1\},\{g_2\}\ldots]$ as the eigenvector matrix associated with $[A]$, where the jth column vector $\{g_j\}$ represents the jth eigenvector. Multiplying Eq. (16) on the left by $[G]^{-1}$, the inverse of the eigenvector matrix, and using the identity matrix $[G][G]^{-1}=I$, an uncoupled vector differential equation for the transformed uncertainty coefficient vector $\hat{\boldsymbol{\gamma}}=[G]^{-1}\boldsymbol{\gamma}$ is obtained,

$$2ik_o\frac{\partial\hat{\boldsymbol{\gamma}}}{\partial r} + \frac{\partial^2\hat{\boldsymbol{\gamma}}}{\partial z^2} + [\Omega]\hat{\boldsymbol{\gamma}} = 0. \quad (18)$$

The diagonal matrix $[\Omega]=[G]^{-1}[A][G]$ contains the eigenvalues $\boldsymbol{\lambda}=\{\lambda_1,\lambda_2,\ldots\}$ of $[A]$; since the coupling matrix is real and symmetric, the eigenvalues are also real. The solution of Eq. (8) for this simple ensemble of environments now reduces to solving Eq. (18) for each transformed uncertainty coefficient $\hat{\gamma}_l(r,z)$ and performing an inverse operation to obtain $\gamma_l(r,z)$. These latter coefficients determine the stochastic pressure distribution through Eq. (9). The lth transformed coefficient satisfies a narrow-angle parabolic equation of the form

$$2ik_o\frac{\partial\hat{\gamma}_l}{\partial r} + \frac{\partial^2\hat{\gamma}_l}{\partial z^2} + \lambda_l\hat{\gamma}_l = 0; \quad l=0,1,2,\ldots, \quad (19)$$

subject to initial and boundary conditions. The solution of Eq. (19) for $\hat{\gamma}_l$ can be obtained by a separation of variables where $\hat{\gamma}_l(r,z)=R(r)Z(z)$. Substituting this product into Eq. (19) leads to a solution of the form

$$\hat{\gamma}_l(r,z) = \hat{C}_l[\hat{A}_l\sin(\varepsilon z) + \hat{B}_l\cos(\varepsilon z)]e^{i\eta r/2k_o}, \quad (20)$$

where $\varepsilon=\sqrt{\lambda_l-\eta}$. The separation constant, η, and $\hat{A}_l,\hat{B}_l,\hat{C}_l$, are determined by the boundary and initial conditions. The boundary conditions can be obtained by starting with the relationship between the stochastic acoustic field and the envelope function. Using Eq. (6b) and Eq. (9), the pressure release surface condition leads to the result $P(r,0) = \sqrt{2/\pi k_o}re^{i(k_o r-(\pi/4))}\Sigma_{l=0}^{\infty}\gamma_l(r,0)He_l(\xi)=0$. For arbitrary ξ, this implies that $\gamma_l(r,0)=0$. Note that the dummy index q has been replaced here by the dummy index l. Similarly,

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Steven I. Finette: Stochastic representation of uncertainty   2573

the rigid bottom condition gives $[\partial \gamma_l(r,z)/\partial z]_{z=D}=0$. The boundary conditions on the transformed coefficients can be obtained by using the corresponding vector forms and multiplying them on the left by $[G]^{-1}$. For arbitrary l,

$$\hat{\gamma}_l(r,0) = 0, \tag{21a}$$

$$\left|\frac{\partial \hat{\gamma}_l(r,z)}{\partial z}\right|_{z=D} = 0. \tag{21b}$$

Applying the constraints specified in Eqs. (21) to Eq. (20) forces $\hat{B}_l$ to vanish. The function ε is restricted to the discrete set of values $(2m-1)(\pi/2D), m=1,2,\ldots,M$, where m is limited to those M integers that satisfy the constraint $\chi = \sin^{-1}(k_z/k) \leq 15°$, consistent with the narrow-angle approximation. The separation constant is determined from ε, and after the depth-dependent modes in Eq. (20) are normalized to unity, one obtains the following result:

$$\hat{\gamma}_l(r,z) = \sqrt{\frac{2}{D}} \sum_{m=1}^{M} C_{ml} \sin\left[(2m-1)\frac{\pi z}{2D}\right]$$

$$\times e^{\{i[\lambda_l - [(2m-1)(\pi/2D)]^2]r\}/2k_o}, \tag{22}$$

where the constant C_{ml} is to be computed from the initial condition and the additional subscript anticipates a dependence on mode number. A Gaussian starter field is considered for the initial condition,[38] so that the envelope evaluated at the range origin is given by

$$\psi(0,z) = \sqrt{k_o} e^{-(k_o^2/2)(z-z_o)^2}. \tag{23}$$

To apply the initial condition, the inverse transform is applied to the vector $\hat{\gamma}: \hat{\gamma}=[G]\gamma$, where the components of $\hat{\gamma}$ are given by Eq. (22). Each γ_l is therefore a linear combination of $\hat{\gamma}_l$, weighted by the eigenvectors of the coupling matrix $[A]$. Equating the polynomial chaos solution [Eq. (6b)] for $\psi(r,z;\theta)$ with Eq. (23) at range $r=0$ yields $\sum_{l=0}^{L}\gamma_l(0,z)He_l(\xi)=\sqrt{k_0}e^{-(k_0^2/2)(z-z_0)^2}$. Explicitly, one obtains

$$\sum_{l=0}^{L}\left[\sum_{u=1}^{L} g_{lu}\left\{\sqrt{\frac{2}{D}}\sum_{m=1}^{M} C_{mu}\sin\left((2m-1)\frac{\pi z}{2D}\right)\right\}\right]He_l(\xi)$$

$$= \sqrt{k_o}e^{-(k_o^2/2)(z-z_o)^2}, \tag{24}$$

where L is the number of terms (i.e., order) retained in the truncated polynomial chaos expansion. Limiting the series representation of the random process to a finite number of terms introduces a truncation error in the estimation of the field moments.[24] The above equation can be solved for the coefficients C_{mu} by first multiplying Eq. (24) by He_r and ensemble averaging both sides. Noting that on the right-hand side $\langle He_r \rangle = \langle He_r He_0 \rangle = \delta_{r0}$,

$$\sum_{u=1}^{L} g_{ru}\sqrt{\frac{2}{D}}\left\{\sum_{m=1}^{M} C_{mu}\sin\left[(2m-1)\frac{\pi z}{2D}\right]\right\}$$

$$= \sqrt{k_o}e^{-(k_o^2/2)(z-z_o)^2}\delta_{r0}, \tag{25}$$

Multiplying Eq. (25) by another member of the orthonormal set of eigenvectors, $\sqrt{2/D}\sin[(2m'-1)(\pi z/2D)]$, and integrating over the waveguide depth yields

$$\sum_{u=1}^{L} g_{ru} C_{mu} = \sqrt{\frac{2k_o}{D}}\left\{\int_0^D \sin\left[(2m-1)\frac{\pi z}{2D}\right]\right.$$

$$\left.\times e^{-(k_o^2/2)(z-z_o)^2}dz\right\}\delta_{r0}. \tag{26}$$

Operating on Eq. (26) with $[G]^{-1}$ gives the coefficients in the form

$$C_{mu} = \sqrt{\frac{2k_o}{D}}a_{u1}\Gamma(m,z_0), \tag{27}$$

where a_{u1} represents the uth element of the inverse of the first eigenvector and $\Gamma(m,z_o)$ is the definite integral in Eq. (26). Substituting Eq. (27) into Eq. (22) gives the explicit solution to Eq. (19) for the uth transformed uncertainty coefficient,

$$\hat{\gamma}_u(r,z) = \frac{2\sqrt{k_o}}{D}a_{u1}\sum_{m=1}^{M}\Gamma(m,z_o)\sin\left[(2m-1)\frac{\pi z}{2D}\right]$$

$$\times e^{\{i[\lambda_u - [(2m-1)(\pi/2D)]^2]r\}/2k_o} \tag{28}$$

Finally, the stochastic pressure field is obtained in the far field using Eq. (9) and Eq. (28),

$$P(r,z;\theta) \approx \sqrt{\frac{2}{\pi k_o r}}e^{i(k_o r - (\pi/4))}$$

$$\times \sum_{q=0}^{Q}\left[\sum_{u=0}^{Q} g_{qu}\hat{\gamma}_u(r,z)\right]He_q(\xi(\theta)), \tag{29}$$

where the quantity in brackets is $\gamma_q(r,z)$ and Q is the chosen truncation order.

C. Comparison of the polynomial chaos solution with Monte Carlo results

A comparison is now made between the analytic solutions for the first two statistical moments obtained from Eqs. (28) and (29), and independent Monte Carlo computations. The latter are based on computing multiple realizations of the acoustic field using the analytic solution of the deterministic parabolic equation given by Eq. (4). A comparison of Eq. (4) with Eq. (19) indicates that separation of variables can also be applied to solve Eq. (4). The resulting solution of the deterministic narrow angle equation, P_{det}, is then given by

$$P_{\text{det}}(r,z) = e^{-i(\pi/4)}\sqrt{\frac{8}{\pi r}}\left(\frac{1}{D}\right)\sum_{m=1}^{M}\Gamma(m,z_o)\sin\left[(2m-1)\frac{\pi z}{2D}\right]$$

$$\times e^{i([k^2-[(2m-1)(\pi z/2D)]^2-k_o^2]/2k_o+k_o)r}. \tag{30}$$

where $k=\sqrt{k_r^2+k_z^2}$ is the total wave number and Eq. (30) was obtained using both the same boundary /initial conditions and narrow-angle constraint on the number of modes that was discussed above for the polynomial chaos solution. The comparison involves the first and second-order moments of the field computed at frequencies of 50 and 300 Hz. The Monte Carlo estimates of the moments are obtained from Eq. (30) in the standard manner, drawing

3000 realizations of the sound-speed field from a Gaussian distribution with $\bar{c}=1500$ m/s and $\sigma=5$ m/s. A source depth of $z_o=40$ m and waveguide depth $D=200$ m are used in the computation. The polynomial chaos estimate of the moments incorporates the same values of $\bar{c}, \sigma, z_o$, and D.

With the Gauss-Hermite basis functionals, acoustic field moments based on the polynomial chaos expansion are computed from Eq. (9), noting that the envelope is given explicitly by the double summation in Eq. (29). The coherent mean field $\langle P(r,z;\theta) \rangle$ is obtained by ensemble averaging both sides of Eq. (9) and using $He_0(\xi)=1$ in conjunction with the orthogonality condition $\langle He_q He_0 \rangle = \delta_{q0}$. The result is

$$\langle P(r,z;\theta) \rangle = \sqrt{\frac{2}{\pi k_o r}} e^{i(k_o r - (\pi/4))} \gamma_0(r,z). \quad (31)$$

The autocovariance, R_{pp}, which includes the field variance as a special case, is defined by the expression

$$R_{pp}(r_1,z_1;r_2,z_2) \equiv \langle [P(r_1,z_1;\theta) - \langle P(r_1,z_1;\theta) \rangle]$$
$$\times [P(r_2,z_2;\theta) - \langle P(r_2,z_2;\theta) \rangle]^* \rangle, \quad (32)$$

where * denotes complex conjugation. Substituting Eq. (9) and Eq. (31) into Eq. (32) and invoking orthogonality of the basis functions gives

$$R_{pp}(r_1,z_1;r_2,z_2) = \frac{2}{\pi k_o \sqrt{r_1 r_2}} e^{ik_o(r_1-r_2)}$$
$$\times \sum_{q=1}^{Q} \gamma_q^*(r_1,z_1) \gamma_q(r_2,z_2) \quad (33)$$

for the autocovariance of the pressure field between points 1 and 2. Note that the sum in Eq. (33) starts at unity rather than zero. The field autovariance for any $(r_1=r_2,z_1=z_2)$ is a special case of Eq. (33) and can be written concisely as

$$R_{pp}(r,z;r,z) = \frac{2}{\pi k_o r} \sum_{q=1}^{Q} |\gamma_q(r,z)|^2. \quad (34)$$

When the ensemble mean in Eq. (31) is negligible, $10 \log_{10} R_{pp}(r,z;r,z)$ gives an unreferenced estimate of coherently averaged transmission loss. Figure 2 shows the magnitude of the coherent mean field over a range of 20 km computed from Eq. (31) for both frequencies. The black curve in each case corresponds to the Monte Carlo estimate of the magnitude of the first moment obtained from 3000 realizations using Eq. (30), and the number of terms Q in the truncated polynomial chaos expansion is $Q=50$ for the 50-Hz case and $Q=400$ for the 300-Hz case. Coherence degradation occurs at a shorter range for the 300-Hz case (containing 21 modes) relative to the 50-Hz case (with three modes). First moment estimates using polynomial chaos expansions are in excellent agreement for both frequencies, with discrepancies for $r>16$ km at 50 Hz and near ranges 8.5 and 18 km at 300 Hz. The autovariance is illustrated in Fig. 3 for both frequencies. The comparison is again very good with $Q=50$ for the 50-Hz case and $Q=400$ for the 300-Hz case, though a discrepancy exists at a range of about 18 km for propagation at both frequencies.The errors may be related to a conver-
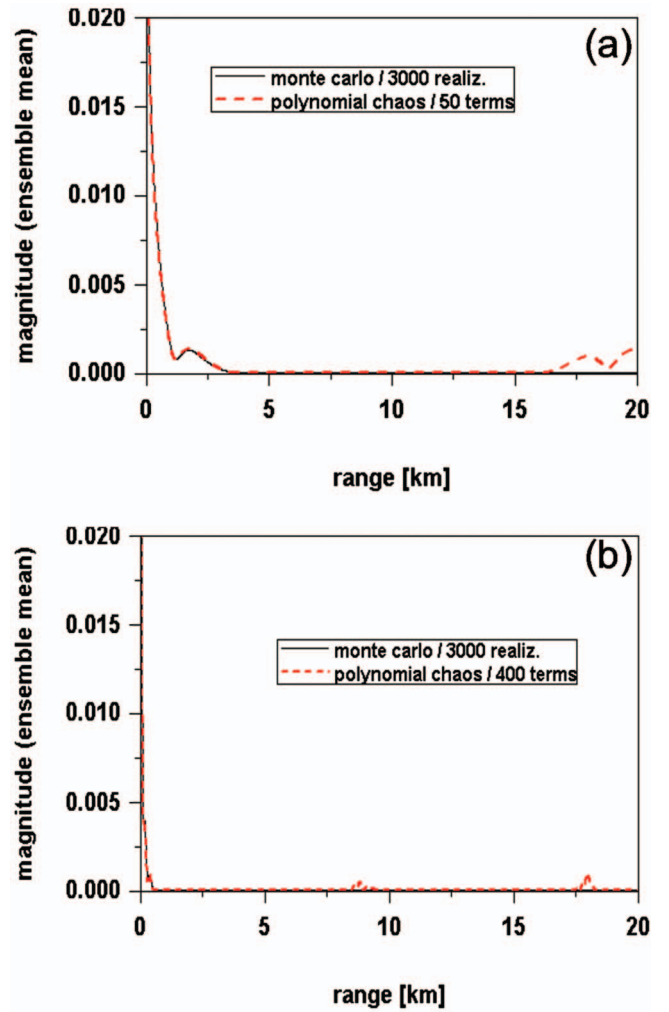
FIG. 2. Comparison of the magnitude of the first moment (coherent mean) of the stochastic pressure field, using polynomial chaos expansions and Monte Carlo sampling: (a) for 3 modes at 50 Hz; (b) for 21 modes at 300 Hz.

gence issue linked to truncation of the series representation in Eq. (29). Some additional comments on this point are given after the results for the autocovariance are presented.

The normalized autocovariance was computed from Eq. (33) for a fixed depth slice where $(r_1=5$ km, $z_1=40$ m, $r,z)$ and is presented in Fig. 4. Polynomial chaos results are again in excellent agreement with those using Monte Carlo-based realizations for 50 Hz, though an additional correlation occurs in the 300-Hz example at a range of about 14 km. Similar results are found for other range/depth choices (not shown). Because the PC expansion can be viewed as a generalized spectral expansion,[11,27] this spurious correlation, identical in form to the correct one located at a range of 5 km, is tentatively interpreted here as a form of wraparound caused by the series truncation to a finite order. Other computations (not illustrated) show that the location of these peaks changes as a function of the number of terms in the expansion. When Q is increased, the distance between the true correlation and the spurious contribution increases. Further work is needed to examine this source of error.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Steven I. Finette: Stochastic representation of uncertainty   2575
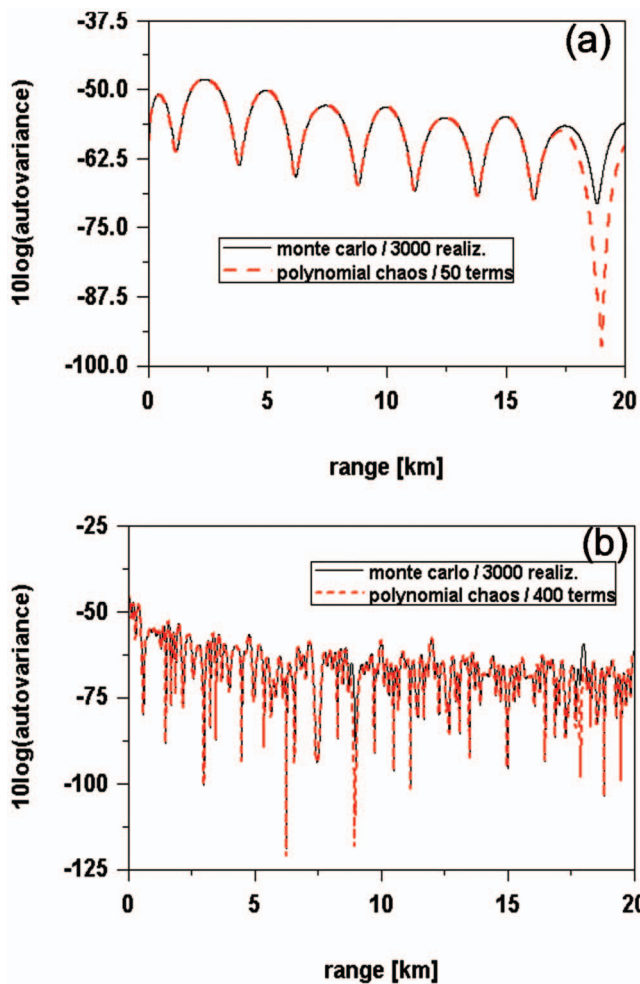
FIG. 3. Comparison of the autovariance of the field using polynomial chaos expansions and Monte Carlo sampling: (a) for 3 modes at 50 Hz; (b) for 21 modes at 300 Hz.

FIG. 4. Comparison of the normalized autocovariance of the field using polynomial chaos expansions and Monte Carlo sampling: (a) for 3 modes at 50 Hz; (b) for 21 modes at 300 Hz.

D. Application of polynomial chaos expansions in realistic environments

It appears that a fairly high-order chaos expansion is necessary to achieve a good comparison with moments independently computed through Monte Carlo sampling, considering the simple structure of the waveguide environment used in the above example. A recent analysis of one-dimensional propagation in unbounded random media[23] indicates that, in the case where the spatial correlation length of the medium is infinite (corresponding to the isospeed case discussed here), the convergence of the PC expansion is non-uniform in range and a significant number of terms are necessary for convergence of the truncated series. However, for media with finite correlation lengths, such as those found in typical ocean waveguides, that analysis predicts that the number of terms necessary for convergence decreases significantly.[23] The convergence rate in both cases depends on both the correlation length and the strength of the sound-speed perturbations. The example considered in Secs. III B and III C involves bounded multipath propagation in a two-dimensional waveguide and, while idealized, has the advantage of being both analytically tractable and illustrative of the general methodology. On the other hand, in light of the

above comments the example would appear to present a particularly difficult case for polynomial chaos expansions due to the infinite correlation length assumed for the medium.

In the remainder of this section a brief outline of an application of polynomial chaos expansions to a more realistic ocean environment is presented, with details left to a future analysis. The problem considers a linear, spatially diffuse internal wave field in a horizontally stratified ocean waveguide. A sound-speed perturbation, $\delta c(x,y,z,t)$, produced by the internal wave field is presumed to be specified by a zero mean random field related to a water particle's random displacement, $\Delta(x,y,z,t)$, from its equilibrium position. This relationship is given by the expression[39]

$$\delta c(x,y,z,t) \cong A N^2(z) \Delta(x,y,z,t). \quad (35)$$

In this equation, A is a constant and $N(z)$ is the (deterministic) buoyancy frequency.

It is assumed that the expansion in Eq. (6a) describing the sound-speed perturbation as a stochastic process is known when solving the forward problem, i.e., computing

the pressure field from Eqs. (8) and (9). It was assumed earlier that δc was characterized by a single random variable ξ. Realistic ocean environments have finite correlation lengths and this leads to a more complex expansion. If the correlation function is known or can be estimated, another form of stochastic expansion may be used to specify δc as a random field in terms of a set of uncorrelated random variables $\{\xi_1, \ldots, \xi_m\}$. The Karhunen-Loeve expansion[27] (also known as principle component analysis or empirical orthogonal function expansion) of the sound-speed perturbation can be written as

$$\delta c(x,y,z,t;\theta) = AN^2(z) \sum_{m=0}^{\infty} \sqrt{\lambda_m} \xi_m(\theta) f_m(x,y,z,t), \quad (36)$$

where λ_m and $f_m(x,y,z,t)$ are the eigenvalues and eigenfunctions, respectively, of the particle displacement correlation function. This function is related to the sound-speed correlation function by

$$\langle \delta c(x,y,z,t) \delta c(x',y',z',t') \rangle$$
$$= A^2 N^2(z) N^2(z') \langle \Delta(x,y,z,t) \Delta(x',y',z,t') \rangle, \quad (37)$$

where θ has been dropped from the notation. Note that both δc and Δ are wide-sense stationary in x,y and t but not in depth. A series of the form in Eq. (36) is optimal in the sense that the mean-square error that results from truncating the sum is minimized for any number of terms retained. The random particle displacement can be expressed explicitly as a linear superposition of internal wave modes with random amplitudes[40]

$$\Delta(x,y,z,t)$$
$$= \sum_{k_x} \sum_{k_y} \sum_{m} G(k_x,k_y,m) W_m(k,z) e^{i(k_x x + k_y y - \omega(k,m)t)}. \quad (38)$$

In this expression, W_m are depth-dependent internal wave eigenmodes, k is the horizontal wave number, and G correspond to a set of zero-mean Gaussian random variables related to the spectral density of the internal wave field, E, of the mth mode by $E(k_x,k_y,m) = \langle G(k_x,k_y,m) G^*(k_x,k_y,m) \rangle$. The dispersion relation between wave number and frequency is given by $\omega(k,m)$. It can be shown[13] that the eigenfunctions f_m in Eq. (36) are given by $W_m(k,z) e^{i(k_x x + k_y y - \omega(k,m)t)}$ and the eigenvalues are given by components of the spectral energy density of the internal wave field, $E(k_x,k_y,m)$. With these identifications, Eq. (36) can be used instead of Eq. (6a) to derive a set of uncertainty equations where the sound-speed perturbations are linked to a linear stochastic internal wave field. The polynomial chaos expansion in Eq. (6b) is still employed as a generic expansion for the envelope field; the resulting series solution for the envelope field is guaranteed to converge for a second-order process, though the rate of convergence would depend on the choice of polynomial chaos basis functionals. Note that while the envelope field is a function of the $\xi_{m'}$ it is not possible to choose a Karhunen-Loeve expansion for the envelope field because the correlation function of the envelope is unknown for the forward problem. The number of terms in the sum needed to specify δc depends on the correlation structure of the sound-speed

distribution; spatial scales of the distribution are embedded in f_m. The eigenvalues weight the importance of the spatial scales included in the uncertainty computation, with the number of terms retained in Eq. (36) determined by the (monotonic) decay rate of the eigenvalues which, in turn, depends on the correlation structure.[11] Two general limiting cases are worth noting: more terms are required in Eq. (36) as the process approaches white noise, while typically only a few terms are necessary for smoothly varying processes, and at the lower limit a spatially uniform field is described by one term. The latter limiting case corresponds to the detailed example discussed earlier. It is expected that a relatively high-dimensional chaos expansion would be required, depending on the spatial frequency band of interest chosen for the internal wave spectrum. On the other hand, based on previous applications of the method it is presumed that a low order is probably sufficient.[11,27,29–31] The resulting uncertainty equations would be solved numerically to obtain the both uncertainty coefficients and the moments.

IV. SUMMARY AND CONCLUSIONS

Simulation-based prediction involves a complex scientific hypothesis in which uncertainty plays a central role. This paper offers a solution to the problem of how statistical characterization of incomplete environmental knowledge can be included within the framework of simulation-based prediction. A dynamical formulation of the problem is presented using a probabilistic approach that leads to stochastic, rather than deterministic wave propagation when uncertainty is included and associated with the sound-speed distribution. Environmental uncertainty linked to sound speed is explicitly coupled to the corresponding uncertainty in the acoustic field and effectively "propagates" through the waveguide along with the field. Measures of the spread of the acoustic field distribution (i.e., statistical moments) are interpreted as quantitative estimates of the effect of environmental uncertainty on acoustic field structure. Polynomial chaos basis functionals are chosen to represent incomplete knowledge of both the sound speed and acoustic fields, and a set of coupled, deterministic partial differential equations are derived for the expansion coefficients describing the stochastic acoustic field. An alternative method for including incomplete environmental information in the dynamics of wave propagation takes a different probabilistic approach, making use of Bayesian inference to describe uncertainty.[41]

The equations are solved exactly for the special case of an isospeed stochastic waveguide within the narrow-angle parabolic approximation and where the variance of the sound-speed fluctuations is small. The sound-speed field is treated as a Gaussian random variable, and Gauss-Hermite functionals are used as a specific example of PC expansion bases. The results for the first two moments are compared to the same moments independently estimated by Monte Carlo analysis of the deterministic narrow-angle parabolic equation, and show that the truncated polynomial chaos expansions reproduce the Monte Carlo estimates of the moments. It should be emphasized that the general method is not limited to either a single source of uncertainty or to the assump-

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Steven I. Finette: Stochastic representation of uncertainty   2577

tions used in the example from Secs. III B and III C. Distributions of sound speed with finite spatial correlation lengths should yield good results with far fewer terms in the polynomial chaos expansion than are needed here. With spatially varying environmental distributions, uncertainty coefficients can be computed numerically from Eq. (8).

Conventional wisdom would say that adding more physics and/or decreasing the grid resolution would lead to better predictions and decreasing uncertainty. However, there are several reasons to question this conclusion. The additional physical parameters or fields may be incompletely specified as mentioned in the Introduction, and this might lead to poorer quality predictions from simulation. Increased spatial and temporal resolution, by itself, may not improve simulation-based prediction unless the physics at those scales is properly incorporated. Uncertainty in the physics at small spatial and temporal scales that is not explicitly modeled can adversely affect the predictions[42,43] by propagating, via the system dynamics, to larger scales of interest. Polynomial chaos expansions may play a significant role in relating uncertainty at different scales, since the expansion coefficients are functions of the space-time correlation structure of the medium.[11] In addition, it is not clear that more "complex" models (i.e., those that include additional physics but do not incorporate uncertainty) have a better predictive capability than simpler models that include uncertainty in their computations. It is possible that embedding a polynomial chaos representation of uncertainty within a relatively simple environmental description of an ocean waveguide may have a better simulation-based predictive capability than a model containing a very detailed environmental representation but without the inclusion of uncertainty in that representation.

The polynomial chaos basis expansion method is one approach for developing a unified theoretical formulation of uncertainty within a strong mathematical framework. It forms a natural foundation for a dynamical formulation of environmental uncertainty and its inclusion into the dynamics of wave propagation, though additional research is needed to address both the numerical issues associated with truncated expansions and applicability of the method to more realistic waveguide environments.

ACKNOWLEDGMENTS

[1]The words knowledge and information are used interchangeably here, e.g., the lack of complete knowledge about a physical system is equivalent to incomplete information or uncertainty concerning the system.

[2]S. Finette, "Propagation of uncertainty for acoustic fields using polynomial chaos expansions," J. Acoust. Soc. Am. **114**, 2461 (2003).

[3]S. Finette, "Embedding uncertainty into ocean acoustic propagation models," J. Acoust. Soc. Am. **117**, 997–1000 (2005).

[4]*Predictability: Quantifying Uncertainty in Models of Complex Phenomena. Special issue of Physica D* **133**, edited by, S. Chen, L. Margolin, and D. Sharp (Elsevier Science, The Netherlands 1999).

[5]D. E. Post and L. G. Votta, "Computational science demands a new paradigm," Phys. Today, 35–41 Jan 2005.

[6]N. Oreskes, K. Shrader-Frechette, and K. Belitz, "Verification, validation, and confirmation of numerical models in the earth sciences," Science **263**, 641–646 (1994).

[7]J. S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2001).

[8]K. Sobczyk, *Stochastic Wave Propagation* (Elsevier, The Netherlands, 1985).

[9]S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics. 4. Wave Propagation Through Random Media.* (Springer, New York, 1989).

[10]T. Y. Hou, H. Kim, B. Rozovskii, and H. Zhou, "Wiener chaos expansions and numerical solutions of randomly forced equations of fluid mechanics," J. Comput. Phys. **216**, 687–706 (2006).

[11]R. Ghanem, "Probabilistic characterization of transport in heterogeneous media," Comput. Methods Appl. Mech. Eng. **158**, 199–220 (1998).

[12]D. Xiu and G. Karniadakis, "The Wiener-Askey polynomial chaos for stochastic differential equations," SIAM J. Sci. Comput. (USA) **25**, 619–644 (2002).

[13]L. B. Dozier and F. D. Tappert, "Statistics of normal mode amplitudes in a random ocean. I. Theory," J. Acoust. Soc. Am. **63**, 353–365 (1978).

[14]L. B. Dozier and F. D. Tappert, "Statistics of normal mode amplitudes in a random ocean. II. Computations," J. Acoust. Soc. Am. **64**, 533–547 (1978).

[15]A. G. Sazontov, "Acoustic coherence in a deep random oceanic waveguide," *in The Formation of Acoustical Fields in Oceanic Waveguides*, edited by. V. Talanov and V. Zverev (Russian Academy of Sciences, Nizhny, Novgorod, 1995), pp. 37–62.

[16]S. Frankenthal and M. J. Beran, "Propagation in random stratified waveguides—A modal-spectral treatment," J. Acoust. Soc. Am. **104**, 3282–3295 (1998).

[17]A. M. Whitman and M. J. Beran, "Pulse propagation of acoustic waves scattered in a three-dimensional channel," Wave Motion **10**, 231–251 (2000).

[18]B. J. Uscinski, "Parabolic moment equations and acoustic propagation through internal waves," Proc. R. Soc. London, Ser. A **372**, 117–148 (1980).

[19]B. J. Uscinski, C. Macaskill, and T. E. Ewart, "Intensity fluctuations. I. Theory," J. Acoust. Soc. Am. **74**, 1474–1483 (1983).

[20]D. B. Creamer, "Scintillating shallow water waveguides," J. Acoust. Soc. Am. **99**, 2825–2838 (1996).

[21]P. Ratilal and N. C. Makris, "Mean and covariance of the forward field propagated through a stratified ocean waveguide with three-dimensional random inhomogeneities," J. Acoust. Soc. Am. **118**, 3532–3559 (2005).

[22]T. Chen, P. Ratilal, and N. C. Makris, "Mean and variance of the forward field propagated through three-dimensional random internal waves in a continental-shelf waveguide," J. Acoust. Soc. Am. **118**, 3560—3574 (2005).

[23]D. B. Creamer, "Comments on using polynomial chaos for modeling uncertainty in acoustic propagation," J. Acoust. Soc. Am. **119**, 1979–1994 (2006).

[24]R. V. Field and M. Grigoriu, "On othe accuracy of the polynomial chaos approximation," Probab. Eng. Mech. **19**, 65–80 (2004).

[25]B. J. Debusschere, H. N. Najm, P. P. Pebay, O. M. Knio, R. G. Ghanem, and O. P. LeMaitre, "Numerical challenges in the use of polynomial chaos representations for stochastic processes," SIAM J. Sci. Comput. (USA) **26**, 698–719 (2004).

[26]S. Hassani, *Mathematical Physics. A Modern Introduction to its Foundations* (Springer, New York, 1999).

[27]R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, New York, 2001).

[28]N. Wiener, "The homogeneous chaos," Am. J. Math. **60**, 897–936 (1938).

[29]O. LeMaitre, O. Knio, H. Najm, and R. Ghanem, "A stochastic projection method for fluid flow I. Basic formulation," J. Comput. Phys. **173**, 481–511 (2001).

[30]D. Xiu, D. Lucor, C.-H. Su, and G. Karniadakis, "Stochastic modeling of flow structure interactions using generalized polynomial chaos," ASME Trans. J. Fluids Eng. **124**, 51–79 (2002).

[31]R. Ghanem and B. Hayek, "Probabilistic modeling of flow over rough terrain," ASME Trans. J. Fluids Eng. **124**, 42–50 (2002).

[32]R. M. Gray and L. D. Davisson, *Random Processes. A Mathematical Approach for Engineers* (Prentice-Hall, Englewood Cliffs, NJ, 1986).

[33]R. Cameron and W. Martin, "The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals," Ann. Math. **48**, 385–392 (1947).

[34]F. D. Tappert, "The parabolic approximation method," *in Wave Propaga-*

tion and Underwater Acoustics, edited by J. B. Keller and J. S. Papadakis, Lecture Notes in Physics, Vol. **70** (Springer, New York, 1977), pp. 224–287.

[35] J. B. Keller, "Wave propagation in random media," Proc. Symp. Appl. Math. **13**, 227–246 (*Amer. Math. Soc.*, Providence, 1962).

[36] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[37] G. D. Manolis and S. Pavlou, "Fundamental solutions for SH-waves in a continuum with large randomness," Eng. Anal. Boundary Elem. **23**, 721–736 (1999).

[38] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).

[39] W. H. Munk and F. Zachariasen, "Sound propagation through a fluctuating stratified ocean: Theory and observation," J. Acoust. Soc. Am. **59**, 818–838 (1976).

[40] S. Flatte, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).

[41] K. R. James and D. R. Dowling, "A probability density function method for acoustic field uncertainty analysis," J. Acoust. Soc. Am. **118**, 2802–2810 (2005).

[42] B. DeVolder, J. Glimm, J. W. Grove, Y. Kang, Y. Lee, K. Pao, D. H. Sharp, and K. Ye, "Uncertainty quantification for multiscale simulations," ASME Trans. J. Fluids Eng. **124**, 29–41 (2002).

[43] J. Glimm, S. Hou, H. Kim, Y. Lee, D. Sharp, K. Ye, and Q. Zou, "Risk management for petroleum reservoir production: A simulation-based study of prediction," Comput. Geosci. **5**, 173–197 (2001).

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Steven I. Finette: Stochastic representation of uncertainty   2579

Scattering from the mixed layer base into the sound shadow

Daniel L. Rudnick[a] and Walter Munk

*Scripps Institution of Oceanography, University of California at San Diego,
La Jolla, California 92093-0225*

Long-range sound transmissions in the ocean are largely controlled by the SOFAR wave guide. The agreement between the measured and predicted arrival patterns is generally excellent, except for an observed extension downward by many hundreds of meters from the lower caustics into the deep sound shadow. The deepening of early arrivals is proposed to be associated with the sharp transition zone marking the lower boundary of the surface mixed layer. This mixed layer base (MLB) is distorted by internal waves. Multiple collisions of the ray-like acoustic transmission with the wavy MLB lead to a mean deepening of the lower caustics of $2\gamma_a^{-1}\sigma^2 \approx 35$ m per collision, where $\gamma_a = (1/C)dC/dz = 1.13 \times 10^{-2}$ km^{-1} is the (fractional) abyssal adiabatic sound speed gradient and $\sigma^2 = 2 \times 10^{-4}$ is the variance in MLB slope. There are typically 20 such collisions in a 1000 km transmission. Monte Carlo numerical experiments yield statistics of ray inclination, range, travel time, and lower turning point. The resulting time front includes a deepening by several hundred meters. The acoustic signatures provide the possibility for monitoring upper ocean processes with abyssal acoustic arrays. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2338813]

I. INTRODUCTION

An acoustic pulse at megameter ranges is typically dispersed over 3–5 s (the SOFARGRAM). The first second sees the arrivals of steep rays that span the ocean from top to bottom. These are followed by rays of diminishing steepness and increasing concentration near the axis, with the final arrivals along the sound axis (the famous SOFAR *finale*). As the insonification lifts off the bottom, the region below the lower turning point (LTP) of refracted rays is expected to lie in the sound shadow. In fact, signals are observed many hundred meters beneath the caustics associated with LTPs. The deep arrivals are an anomaly to an otherwise excellent agreement with geometric optics (except for the *finale*). Diffraction cannot account for the observed penetration into the ray shadow (Worcester, 1977; Dziecuich *et al.* 2004).

The upper ocean is associated with two marked boundaries: one at the ocean surface, the other at the base of the mixed layer (MLB). Recent observations have revealed a remarkable sharpness of the MLB, capable of the reflection of low-frequency acoustics. Multiple reflections from a MLB distorted by internal waves may account for the observed shadow penetration.

II. PENETRATION INTO THE ABYSSAL SHADOW ZONE

Since the early days of the U.S. Navy Sound Surveillance System (SOSUS) it has been known that late "forbidden" rays are in fact recorded on deep receiving arrays, but the published accounts are very few. Spiesberger and Tappert (1996) refer to "shadow-zone arrivals" in their 3700 km transmissions from Hawaii toward the mainland and spoke of the "the demise of the idea of axially trapped energy."

Dushaw *et al.* (1999) refer to an "interesting, and perhaps new, type of 'ray' arrival which appears to occur well into the shadow zone of the predicted arrival…; the apparent stability of these arrivals, together with the detection of similar arrivals at several other deep arrays in both the Atlantic and Pacific oceans, rules out bottom interaction as the origin of these arrivals. To date, no known mechanism, e.g., diffraction leakage from the caustics or diffusion of acoustic energy by internal wave scattering, can explain the extreme depth diffusion of acoustic energy that must be occurring."

Figure 1 shows results of transmissions from an ATOC source on Pioneer Seamount toward the bottom-mounted horizontal SOSUS line array **n** (Dushaw *et al.*, 1999). For all but the earliest arrivals, the receiver depth is beneath the predicted lower turning points of the rays, yet the signal is clearly received. The authors note that the deep arrivals are stable, and that they occur very near the time of the predicted caustics above. The same result is found for transmissions to other SOSUS arrays, all in the range of one or two megameters. We will discuss the deep arrivals in terms of *reflection* from a wavy MLB. This serves as a simple model to generate the required statistics, but needs to be generalized to account for the later *refracted* arrivals with upper turning points well below the MLB.

III. OBSERVED SOUND SPEED AND THE MIXED LAYER BASE IN THE NORTH PACIFIC

The North Pacific Acoustic Laboratory (NPAL) Spice04 experiment had the objective of quantifying the effect of upper-ocean sound speed fluctuations on long-range acoustic propagation. In May-June 2004, the NPAL group deployed three acoustic moorings along a 1000-km section in the North Pacific, with end points at 34°53'N, 148°25'W and 33°25'N, 137°41'W. Subsequent cruises in September-October 2004 and March-April 2005 afforded opportunities

---

[a]Electronic mail: drudnick@ucsd.edu
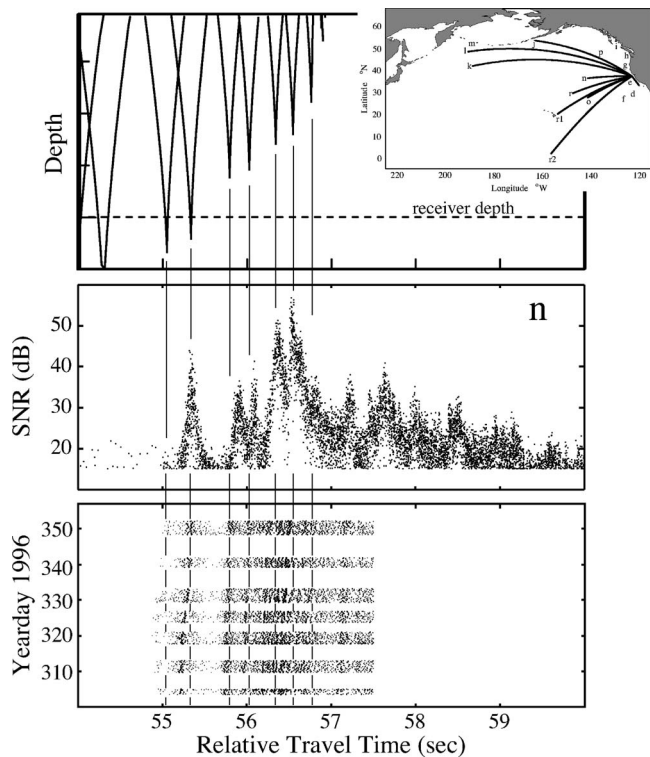
FIG. 1. Predicted time fronts (top) and measured intensities for transmissions from the ATOC source on Pioneer Seamount to SOSUS stations **n** and **o** (adapted from Figs. 3 and 6 of Dushaw *et al.*, 1999). The lowest panel shows the recorded intensities of individual transmissions during year days 300 to 360 in 1996 as a function of transmission time (relative units). The central panel shows the corresponding signal to noise ratios based on all transmissions between January 1996 and March 1997.

to measure hydrography along the section. Observations of temperature, salinity, and pressure were obtained during the first two cruises using the newly developed Underway CTD (Rudnick *et al.*, 2004), a device designed to collect profiles while under way so as not to require ship time or to interfere with other operations. The last cruise was dedicated to measuring fine-scale hydrography using a SeaSoar.

Representative profiles from these three cruises show seasonal changes in the mixed layer and its base. The mixed layer is at its deepest, near 100 m, in late winter or early spring (Fig. 2: April profile). The mixed layer shoals rapidly in the spring, due to increased surface heating, and is near its shallowest in late spring or early summer (see June). The mixed layer warms throughout the summer to its warmest, least dense, state in the fall (see September). The MLB is defined by remarkably sharp gradients in temperature, and therefore sound speed, in a vertical span of about 10–20 m. The net change in sound speed across the MLB can be larger than 10 m/s (see September). This change is equivalent to about 1/5 the variation in sound speed over the entire ocean depth, and five times the change through the mixed layer above the MLB.

For the purpose of this paper, the ocean sound channel is modeled analytically (the "canonical" profile) with a discontinuous transition at 70 m depth at the MLB (Fig. 2). The sound speed in the mixed layer is at the adiabatic gradient, with no net change in heat content relative to the canonical profile above the MLB. The resulting sound speed change at
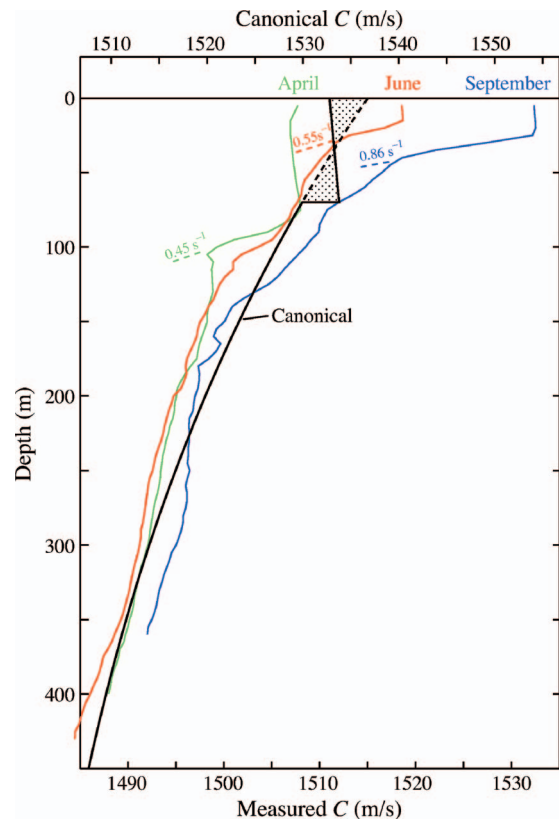
FIG. 2. (Color) Sound speed profiles (lower scale) from April 2005, June 2004, and September 2004, representative of spring, summer, and fall in the northeast Pacific. The sharp gradient at the bottom of the mixed layer first forms in June and deepens to about 100 m at the end of the winter storms (a weak deep residual from the previous year is evident in the June profile). The black line (upper scale) is a fit of the canonical profile to the situation in spring. In this paper the mixed layer is modeled by a discontinuity at 70 m depth with no net exchange of heat.

the MLB is about 4 m/s. Observations are consistent with velocity jumps up to 15 m/s, only 1% of the sound speed but one-third of the total variation in the ocean column (ignoring the associated density jumps). Even so, the jump is of course dwarfed by the 80% velocity contrast and the 800:1 density contrast at the sea-air boundary.

Internal waves at the mixed layer base. A SeaSoar tow during the March-April 2005 cruise focused on the MLB (Fig. 3). The nominal depth range 50–150 m was profiled initially, completing a cycle in less than 3 min, resulting in a horizontal resolution of at least 720 m at the tow speed of 4 m/s. As the tow proceeded to the east, the mixed layer deepened, from near 80 to near 140 m, so the SeaSoar profiling range was lowered. The MLB is clear in the figure as a transition from relatively well-mixed water above to stratification below. Statistics of MLB displacement and slope were calculated over wavelengths from the length of the tow (1000 km) to the Nyquist wavelength of 1440 m. Root-mean-square MLB displacement and slope were rms ζ =8.6 m and 0.0034 rad, respectively. The spectrum of slope (Fig. 4) was white out to the Nyquist frequency at a level of about $F_m(k_1)=0.02$ rad^2/cpm (cycles per meters). Previous SeaSoar tows at constant pressure and along isopycnals reveal that the slope spectrum is white at the highest observed
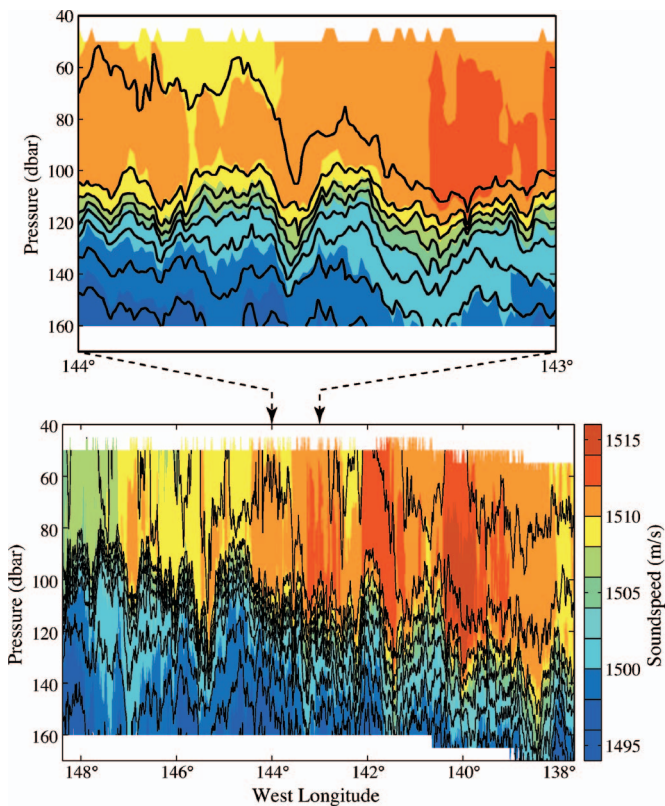
FIG. 3. (Color) Sound speed as color from April 2005 in the northeast Pacific. Black lines are isopycnals (contour interval 0.1 kg/m^3). The total range is almost exactly 1000 km. The transition at the bottom of the mixed layer is clearly indicated. SeaSoar cycled only over the 100 m surrounding the mixed-layer base, yielding a horizontal resolution of 720 m. Mean depth of the 25.36 isopycnal (corresponding roughly to the 1499 m/s sound speed contour) is 115 m. The rms displacement (after removing the trend) is 8.6 m, rms slope is 0.0034. Because the displacement spectrum has a −2 slope, the slope spectrum is white. Thus the rms slope is strongly dependent on horizontal resolution (Nyquist wave number).

wave numbers of order 0.1 m^{-1}. (Hodges and Rudnick, 2006).

The processes causing the MLB to heave are likely mesoscale eddies and internal waves. It is worthwhile to compare the observed statistics to those obtainable through the
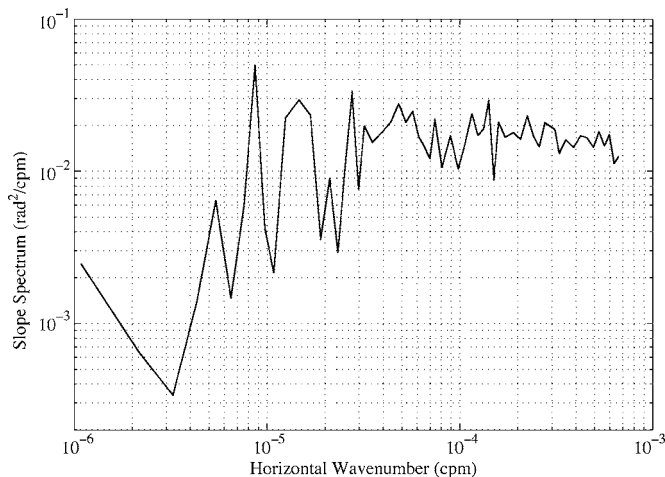
FIG. 4. Slope spectrum of the 25.36 isopycnal. The high wave number termination is determined by the sampling resolution.

Garrett–Munk (GM) spectrum. The observed rms MLB displacement of 8.6 m compares favorably to the GM value of 7.6 m. The GM slope spectrum is white, as is the observed spectrum, but the observed spectral density of 0.02 rad^2/cpm is well below the GM value of 0.08 rad^2/cpm. Shortcomings in the GM model are well known, and the rapidly changing stratification near the MLB is a particular challenge. The details of projecting the traditional GM displacement spectrum to the slope spectrum are cumbersome and uninteresting (Appendix A).

Short internal waves with lengths comparable to the acoustic wavelength are associated with Bragg scatter with scattering coefficients vastly smaller than those associated with specular reflections (Brekhovskikh and Lysanov, 1991 Chap. 9). For a very rough estimate of the wave number cut-off appropriate to specular reflection, consider the dimension of a "ray tube" with Fresnel scaling $\sqrt{\lambda R'}$ where λ is the acoustic wavelength and R' a suitable length scale. In a homogeneous ocean this is the total range r; in a periodic wave guide, R' is related to the ray loop range R. For the canonical sound channel, Flatté et al. (1979, Chap. 9) demonstrate that R' is a small fraction of the ray loop range. Taking $R' = R/10 = 5$ km, and $\lambda_{ac} = 10$ m (150 Hz), the resulting Fresnel scale is 220 m, as compared to the 100 m cut-off adopted in this paper. Dzieciuch (2006) finds that the acoustic transmissions are insensitive to internal waves with vertical scales of less than the acoustic wavelength λ_{ac}. The associated horizontal scale is larger by at least an order of magnitude. Accordingly, for $\lambda_{ac} = 10$ m (150 Hz) we adopt a cutoff wavelength of 100 m, or $k_1^* = 0.01$ cpm. Assuming that the spectrum remains white at the level observed to a cutoff wave number k_1^* cpm, the slope variance is $\sigma^2 = F_m(k_1) \times k_1^* = 0.02 \times 0.01 = 2 \times 10^{-4}$ rad^2 (0.8° rms). The reader will appreciate the uncertainty of this estimate associated with the uncertainty in the termination of the slope spectrum.

A quite different consideration has been suggested by Klymak and Moum (2006): a transition near 10^{-2} cpm from internal waves to turbulence. This implies a failure of the discrete IW dispersion $k(j, \omega)$ at the higher wave numbers, where "the weak selective interactions" of the internal waves are replaced by "strong promiscuous interactions" of the turbulent Fourier components (Phillips, 1966). But if such a transition occurs, it is not clear how it would lead to a marked reduction in the reflectivity of a turbulent versus a wave distorted MLB. One may hope that the acoustic measurements can shed some light upon whether the termination is governed by the acoustic frequency or determined by some intrinsic properties of the ocean dynamics.

IV. STATISTICS OF A SINGLE BOUNCE

A. Refraction versus reflection

Figure 5 sketches the perturbations associated with the steep gradients in the bottom boundary of the mixed layer. The MLB gradients (red) are typically an order of magnitude above those associated with the canonical gradients (black) at that depth. Accordingly, the loop AB_2C_2 is shortened relative to the canonical AB_1C_1. For a given angle of incidence, the loop range varies inversely with the gradient dC/dz and
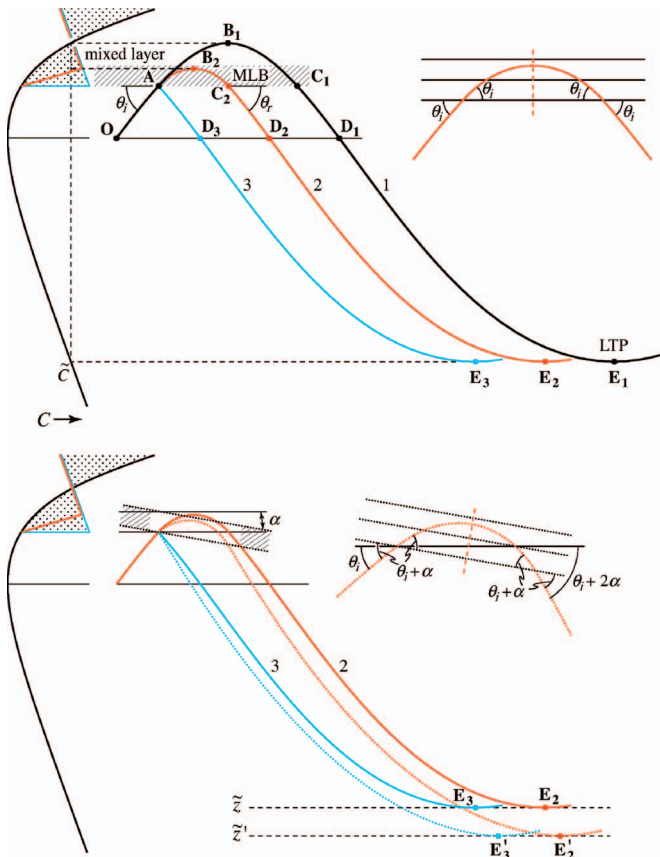
FIG. 5. (Color) Modification of ray geometry by mixed layer base (MLB). The canonical profile (black) is modified by a mixed layer with steep lower boundary (red) and discontinuous boundary (blue). For a level MLB (top) there is no change in the depth of the lower turning point (LTP). For a downward tilting MLB with angle α, both the refracted and reflected rays are steepened by 2α and the LTPs are equally deepened.

so typically $\mathbf{AC_2} \ll \mathbf{AC_1}$. Ray curvature is proportional to sound speed gradient, and as $dC/dz \to \infty$, $\mathbf{AC_2} \to 0$: the downward refracted ray approaches the configuration of the reflected ray (blue). The turning sound speed and depth of the LTP are unchanged; they are specified by the incident inclination θ_i and sound speed C (or slowness $S = 1/C$) at the MLB (assumed level), according to Snell's law

$$S_{\text{MLB}} \cos \theta_i = \widetilde{S}, \tag{1}$$

where S_{MLB} is the slowness at the MLB and $\widetilde{S}$ is turning slowness.

But there is a change in turning depth for an inclined MLB. For a downward sloping MLB (as shown in the following), the *reflected* (blue) ray is steeper than the incident ray by twice the MLB inclination,

$$\theta_r = -\theta_i + 2\alpha. \tag{2}$$

The same "law of reflection" holds for a downward *refracted* ray (red). For simplicity the following discussion is in terms of a reflection at the MLB boundary.

There is total reflection for $C_{\text{mix}} \geqslant \widetilde{C} = 1/\widetilde{S}$ (the refracted ray turns within the layer). For $C_{\text{mix}} < \widetilde{C}$ there is partial reflection (Brekhovskikh and Lysanov, 1991, 3.1.12, neglecting the density contrast)
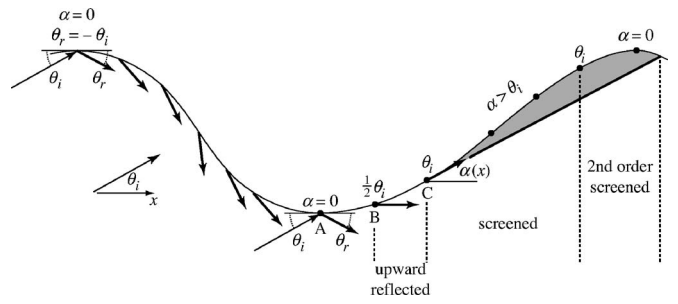
FIG. 6. Ray reflection from a MLB tilted by internal waves with inclination $\alpha(x)$. At the crest and trough ($\alpha=0$) the angles of incidence and reflection are of equal magnitude, $\theta_r = -\theta_i$. Reflections are steepened to the left of the trough (A), and flattened to the right. Reflections are upward for $\frac{1}{2}\theta_i \leqslant \alpha_i (B \to C)$. The boundary is screened for $\alpha > \theta_i$.

$$R = \frac{\sin \theta - \sqrt{n^2 - \cos^2 \theta}}{\sin \theta + \sqrt{n^2 - \cos^2 \theta}}, \quad n = \frac{S_{\text{mix}}}{S_{\text{MLB}}} = \frac{C_{\text{MLB}}}{C_{\text{mix}}} \leqslant 1 \tag{3}$$

and $R=1$ for $n>1$, with R^2 designating the fraction of reflected energy. The subscripts "MLB" and "mix" refer to the bottom and top of the shaded transition layer. For the canonical profile in Fig. 2 with 4 m/s difference between the sound speed above and beneath the transition, the critical angle is 4.1°: only very flat rays are totally reflected. Observations are consistent with velocity jumps up to 15 m/s, corresponding to 8° of critical incidence. Later, we consider a glancing ray of angle 0.6° and a steep ray of angle 8.9°. Assuming $n=1530/1540=0.9935$ consistent with observations, the glancing ray is totally reflected and the steep ray has $R=0.19$.

B. Screening

Incident rays on a wavy reflecting boundary will see only that part of the boundary that is inclined toward the ray. For an incident ray of inclination θ_i the part of the boundary whose inclination $\alpha > \theta_i$ is said to be "screening" (Beckmann and Spizzichino, 1987; the designation "shadowed," sometimes used in the literature, is here reserved to the abyssal acoustic shadow). "Second-order screening," which shadows some of the boundary with $\alpha < \theta_i$ (Fig. 6), is ignored.

Inclinations at the MLB are reasonably well represented by a Gaussian probability density function

$$p(\alpha) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{\alpha^2}{2\sigma^2}\right), \tag{4}$$

where σ is the standard deviation of MLB inclination. The conditional PDF of illuminated inclination, given an incident ray of inclination θ_i, is taken as

$$p_{ill}(\alpha | \theta_i) = \begin{cases} \dfrac{p(\alpha)}{D(\theta_i)}, & \alpha < \theta_i \\ 0, & \alpha \geqslant \theta_i \end{cases} \tag{5}$$

where $D(\theta_i)$ is the Gaussian cumulative distribution function (CDF)

$$\langle(\alpha - \langle\alpha\rangle)^2\rangle = \langle\alpha^2\rangle - \langle\alpha\rangle^2 = \sigma^2 + \theta_i\langle\alpha\rangle - \langle\alpha^2\rangle. \quad (10)$$

But the variance is not a good indicator of the spread because of the large asymmetry between positive and negative α, and it is best to refer to the PDF in Fig. 7.

Upward reflections off the MLB are possible, and all subsequent statistics include these relatively few upward reflections. An alternate approach would be to assume that the upward reflections soon encounter the MLB again, and are reflected downward. A statistical model with only downward reflections is achievable using PDF (5) with θ_i on the right-hand side replaced by $\frac{1}{2}\theta_i$. The results from such a model are qualitatively similar to those shown in this paper.

C. Steepening

Rays incident on the MLB with inclination θ_i are reflected (refracted) with inclination

$$\theta_r = -\theta_i + 2\alpha. \quad (11)$$

The change in inclination per bounce is

$$\delta\theta_r = 2\alpha \quad (12)$$

with a mean negative value of $2\langle\alpha\rangle$ [Eq. (8)]. All other statistics follow from the previous discussion (allowing for a factor 2). Though the *expected* value $\langle\delta\theta_r\rangle$ is negative, e.g., steepening (though diminishing rapidly with increasing θ_i), there is significant probability of positive $\delta\theta_r$ (flattening).

D. Deepening

Deepening is associated with the change in turning slowness $\delta\widetilde{S}$. From Snell's law

$$\widetilde{S} = S\cos\theta_r \quad (13)$$

and accordingly

$$\delta\widetilde{S}/S = \delta\cos\theta_r = \cos(-\theta_i + 2\alpha) - \cos(-\theta_i)$$
$$= 2\sin(\theta_i - \alpha)\sin\alpha \approx 2(\theta_i - \alpha)\alpha = 2\beta, \quad \beta \equiv \theta_i\alpha - \alpha^2,$$
$$(14)$$

assuming small angles. The expected value is

$$\langle\beta\rangle = \theta_i\langle\alpha\rangle - \langle\alpha^2\rangle = -\sigma^2. \quad (15)$$

The variance is

$$\langle\beta^2\rangle - \langle\beta\rangle^2 = 2\sigma^4 - \sigma^2\theta_i\langle\alpha\rangle + \sigma^2\theta_i^2. \quad (16)$$

Beneath 2 km depth, the gradient in sound slowness $S(z) = 1/C(z)$ approaches the adiabatic value

$$\gamma_a = (1/S)dS/dz = 0.0113 \text{ km}^{-1}. \quad (17)$$

Hence the turning depth $\widetilde{z}$ changes with turning slowness $\widetilde{S}(z)$ according to

$$\delta\widetilde{z} = \gamma_a^{-1}\delta\widetilde{S}/S \approx 2\gamma_a^{-1}\beta. \quad (18)$$

Remarkably, the expected value $\langle\beta\rangle = -\sigma^2$ is independent of θ_i, and this makes

FIG. 7. (Color) (a) Probability density function of slope α as a function of the incident ray inclination θ_i, scaled by rms MLB inclination σ. The green curve shows the upper termination of the illuminated α. The expected value $\langle\alpha\rangle$ increases with increasing θ_i from $-\sqrt{2/\pi}\sigma$ to 0 (red). (b) Probability density function of $\beta/\sigma^2 = \delta\cos\theta/2\sigma^2$ with an expected value of -1 (red). The maximum value of β is shown by the green curve.

$$D(\theta_i) = \int_{-\infty}^{\theta_i} p(\alpha)d\alpha = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\theta_i}{\sqrt{2}\sigma}\right)\right]. \quad (6)$$

The normalization by the Gaussian CDF ensures that

$$\int_{-\infty}^{\infty} p_{ill}(\alpha|\theta_i)d\alpha = 1. \quad (7)$$

The distribution of illuminated MLB inclination is thus a clipped, normalized Gaussian (Fig. 7).

The illuminated mean and mean-square MLB inclinations are

$$\langle\alpha\rangle = -\sigma^2 p(\theta_i)/D(\theta_i), \quad \langle\alpha^2\rangle = \sigma^2 + \theta_i\langle\alpha\rangle, \quad (8)$$

and the recurrence relation is

$$\langle\alpha^n\rangle = (n-1)\sigma^2\langle\alpha^{n-2}\rangle + \theta_i^{n-1}\langle\alpha\rangle. \quad (9)$$

The variance is

$$\langle \delta \widetilde{z} \rangle = -2\gamma_a^{-1}\sigma^2 = -35 \text{ m} \qquad (19)$$

a convenient measure of deepening per bounce. Here again the large asymmetry suggests a direct reference to the PDF of β [Fig. 7(b)]:

$$p_{ill}(\beta|\theta_i) = \frac{p_{ill}(\alpha^+|\theta_i) + p_{ill}(\alpha^-|\theta_i)}{\sqrt{(\theta_i^2 - 4\beta)}}, \quad \beta < \frac{\theta_i^2}{4},$$

$$\alpha^\pm = \tfrac{1}{2}(\theta_i \pm \sqrt{\theta_i^2 - 4\beta}). \qquad (20)$$

(Care has been taken to allow for the two roots $\alpha^\pm$ to $\beta = \theta_i \alpha - \alpha^2$.) The expected $\langle \beta \rangle = -\sigma^2$ and so $\langle \widetilde{z} \rangle$ is negative (deepening), but there is a significant chance for $\beta > 0$ (raising the caustic) and for occasional $\beta \ll -\sigma^2$ (large deepening).

E. Stretching

Rays in the ocean waveguide (the ocean "Sound Channel") oscillate about the sound axis between upper and lower turning points at $z^\pm(\widetilde{S})$. Reflections at the undulating MLB lead to perturbations in turning slowness $\delta\widetilde{S}$ and turning depths $\delta\widetilde{z}^\pm$. The associated perturbations in the length R and period T of the ray loops are conveniently given in terms of the action variable $A(\widetilde{S})$ and its derivatives $A' = dA/d\widetilde{S}, A'' = d^2A/d\widetilde{S}^2$ (Appendix B). Expanding $A(\widetilde{S})$ in a Taylor series, the deepening of the lower caustic and the stretching (in range and time) of the loops per bounce are given by

$$\delta\widetilde{z}^- = (\gamma_a S)^{-1}\delta\widetilde{S}, \quad \delta R = -A''\delta\widetilde{S}, \quad \delta T = -A''\widetilde{S}\delta\widetilde{S}, \qquad (21)$$

with $\delta\widetilde{S} = S\delta\cos\theta$. The PDFs of all three variables are derivable from the distribution of $\delta\cos\theta(=2\beta)$. The bias for negative $\delta\cos\theta$ is a property of the cosine function and independent of the bias for negative $\delta\theta$ $(=\tfrac{1}{2}\delta\alpha)$ due to screening. The expected $\langle\delta\cos\theta\rangle = -2\sigma^2$ applies even if there is no screening. For positive A'', deepening is associated with stretching.

V. STATISTICS OF MULTIPLE BOUNCES

A. Reflected angle and its cosine

Long-range acoustic propagation involves multiple bounces off the MLB. Screening will lead to a mean steepening of reflected rays. Also, a growth in the variance of reflected angle θ_r will occur as the repeated bounces are essentially a random walk. The quantity of interest is the cosine of the reflected angle,

$$\cos\theta_r = 1 - \tfrac{1}{2}\theta_r^2 + \cdots. \qquad (22)$$

Therefore the mean value of the cosine will decrease as the variance of reflected angle grows. Through Snell's law this results in a decrease in turning slowness and a deepening of the lower turning point.

The PDF of α is a function of θ_i, and for multiple reflections, θ_i is itself a random variable as it is a sum of a series of α's. So the PDF of θ_i after N bounces is not easy to calculate analytically (Appendix C). However, it is relatively straightforward to perform a Monte Carlo simulation
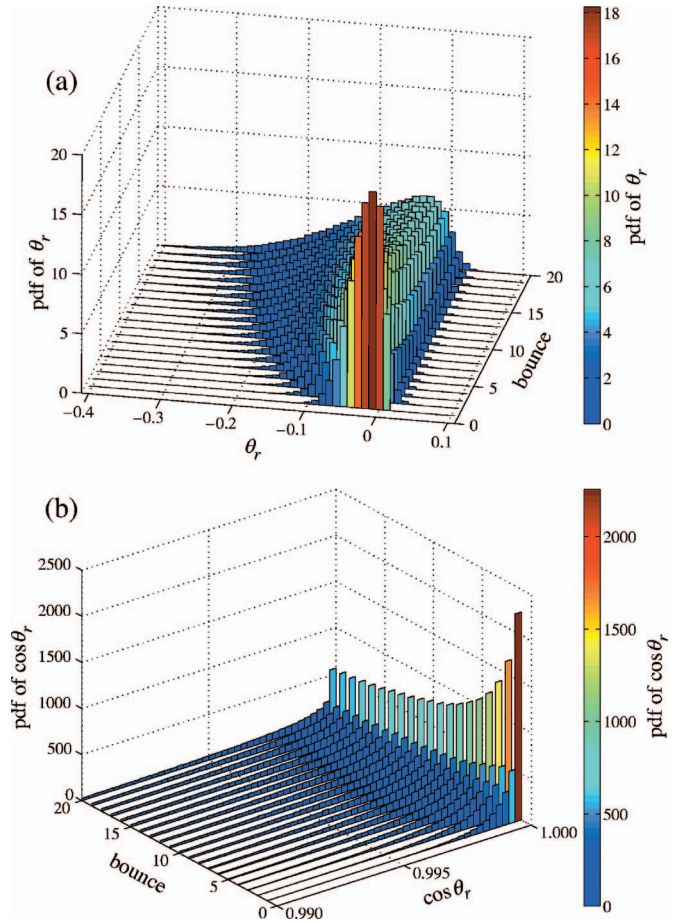
FIG. 8. (Color) Probability density function for θ_r and $\cos\theta_r$ as a function of bounce for glancing initial incidence $\theta_0 = 0.0104$ (0.6°).

of the process using the distribution (5) and reflection law (11) iteratively from an initial MLB inclination θ_0 so $\theta^{(n+1)} = \theta_r^{(n)} + 2\alpha^{(n)}$ with the distribution of $\alpha^{(n)}$ a function of $\theta_r^{(n)}$. Statistics were calculated using 100 000 realizations of 20 consecutive bounces, with the variance of MLB inclination as observed, $\sigma^2 = 2 \times 10^{-4}$. Two choices are considered, $\theta_0 = 0.0104$ rad and $\theta_0 = 0.1555$ rad, representative of angles comparable to, and much larger than, σ.

The glancing initial angle is perhaps the best to investigate first, as it will be most affected by screening, and the assumption of a reflecting MLB is most valid. Results are shown for $\theta_0 = 0.0104$ rad as a sequence of PDFs for each of 20 consecutive bounces [Fig. 8(a)]. The PDF for one bounce is the same as that derived analytically (5). As the number of bounces increases, the likelihood of large negative θ_r increases, due to screening and the increase in variance of a random walk. Because upward reflections are allowed, the PDF leaks over to positive θ_r. The mode of the PDF moves slightly toward steeper angles as the number of bounces increases. The mean reflected angle [Fig. 9(a)] steepens as a function of bounce to a value of -0.1008 rad at 20 bounces. As the mode remains relatively small, this increase in the mean is a result of the growing tail toward large negative angles. The true mean from the Monte Carlo calculation can be compared to the mean [Fig. 9(a)] implied by iterative application of the one-bounce mean (8) and the reflection
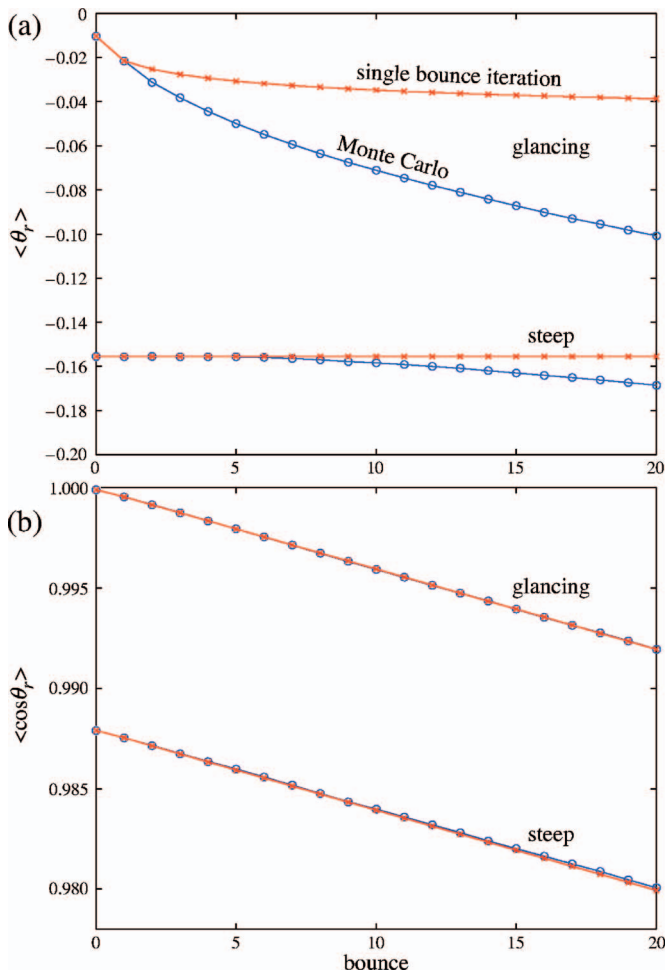
FIG. 9. (Color) Expected θ_r and $\cos \theta_r$ as a function of bounce for glancing incidence $\theta_0 = 0.0104$ (0.6°) and steep incidence $\theta_0 = 0.1555$ (8.9°). Monte Carlo confirms the linear decrease of $\langle \cos \theta_r \rangle$.

formula (11). The true mean increases more rapidly as the sequence of inclinations are serially correlated, with the distribution of each bounce depending on the value of the previous bounce.

The quantity of interest is the cosine of θ_r, which has a PDF [Fig. 8(b)] with a very strong mode at 1. This mode is an elementary property of a function of a random variable whose slope approaches zero. The mode decreases in magnitude with increasing bounces in favor of a long tail toward smaller values of the cosine. This tail is responsible for the decrease in mean cosine [Fig. 9(b)] in the presence of the still large mode at 1. The mean change in the cosine is constant and independent of incident angle for a single bounce (15), so the mean change for multiple bounces should also be constant. This theoretical result agrees nearly perfectly with the Monte Carlo calculation.

A Monte Carlo calculation with a steeper initial angle of 0.1555 rad yields PDFs different in detail, but with the same general features as that for $\theta_0 = 0.0104$ rad after 20 bounces. Growing tails are responsible for changes in mean values of the reflected angle and its cosine, while the large mode at a cosine of 1 remains. The mean reflected angle [Fig. 9(a)] remains relatively constant over the first five bounces, after which it gradually becomes steeper. As the initial angle is

much greater than rms MLB inclination σ, the PDF of α is essentially Gaussian with little screening. The iterative application of Eqs. (8) and (11) supports this interpretation as it predicts virtually no increase in reflected angle. In fact, the reflected angle does steepen because the increasing variance in θ_r eventually creates angles near zero that are affected by screening. The cosine of θ_r [Fig. 9(b)] displays the expected nearly linear decrease with a slight deviation from theory gradually setting in at large bounces. This effect is likely due to the incipient failure of the small angle approximation leading to (15).

B. Range, time, and lower turning point

Examination of the acoustic rays associated with the sequence of angles of the Monte Carlo calculations requires a characterization of the sound channel (Munk *et al.*, 1995, 2.18).[1] The initial angle of 0.0104 rad, used as an example of a glancing ray (Table I), has a total range of 1000 km after 20 bounces off a flat MLB of depth 70 m. The steep ray, initial angle 0.1555 rad, also has a total range of 1000 km.

For the glancing ray, the PDF of loop range [Fig. 10(a)] after one bounce has a mode near the initial loop range of 50.91 km. After three bounces a mode is established at 49.42 km (the shortest possible range, with angle 0.0803 rad), and a tail toward longer range begins to grow. Mean range (Fig. 11) decreases over the first 5 bounces as the mode at the shortest range is established. Growth in mean range then occurs as the tail in the PDF expands. It may, at first, seem curious that the range would shorten while θ_r grows and $\cos(\theta_r)$ decreases. This shortening is understandable as upper loop length decreases with increasing angle up to the angle of minimum loop length, while lower loop length increases monotonically as the angle steepens from zero. The PDF of total range after 20 bounces [Fig. 10(b)] has a mode at 984 km, and a long tail so that the mean total range is 991 km. Even while the mean total range for a wavy MLB is shorter than the total range for a flat MLB, the mean lower turning depth after 20 bounces of 3.56 km is deeper than the initial lower turning depth of 2.94 km. Mean total travel time is 659.88 s compared to 665.72 s for a flat MLB.

The PDF of loop time is similar to that for range because variations in sound speed relative to its absolute value are small. The PDF of lower turning depth is similar to that of $\cos \theta_r$ through Snells law, and because the abyssal gradient of sound speed is nearly uniform. To save space the PDFs of time and lower turning depth are omitted.

The mean range for the steep ray increases monotonically with bounce (Fig. 11). The result is that the total range for the steep ray (1033 km) is more than 40 km longer than that for the glancing ray, with a suitably longer total travel time (686.63 s). The lower turning depth increases from an initial value of 3.88 to a mean of 4.48 km after 20 bounces. So the steep ray has a similar deepening to that of the glancing ray, as expected from the result of the mean change in cosine, but this lower turning point is significantly farther from the source.

TABLE I. Specifications for rays along the time fronts in Fig. 12. Listed are turning slowness $\widetilde{S}$, inclination at the MLB θ_{MLB}, axial inclination θ_{axial}, loop range R, loop time T, upper turning point $\widetilde{z}^+$, lower turning point $\widetilde{z}^-$, and total travel time t. Points along the refracted, glancing MLB reflected, and steep MLB reflected time fronts are listed. The specifications in column C include the reflected glancing and steep rays whose statistics are discussed in the text.

Parameter	+39			+40 and +39	+40	
	A	B	X	C	D	E
Refracted, no mixed layer						
$\widetilde{S}$ (s/km)	0.6542	0.6547	0.6552	0.6558	0.6569	0.6574
θ_{axial} (deg)	11.1	10.9	10.6	10.3	9.81	9.57
R (km)	52.24	51.84	51.38	50.90	50	49.62
T (s)	33.02	33.21	34.20	33.88	33.29	33.05
$\widetilde{z}^+$ (m)	−32	−49	−70	−92	−133	−151
$\widetilde{z}^-$ m	−3063	−3003	−2933	−2860	−2723	−2665
$t-665$ (s)	0.447	0.545	0.696	0.719	0.873	0.951
MLB reflected, glancing						
$\widetilde{S}$ (s/km)			0.655 24	0.655 20	0.6548	0.6544
θ_{MLB} (deg)			0	0.60	2.1	2.9
θ_{axial} (deg)			10.62	10.64	10.8	11.0
R (km)			51.38	50.91	50.00	49.69
T (s)			34.20	33.89	33.30	33.09
$\widetilde{z}^+$ (m)			−70	−70	−70	−70
$\widetilde{z}^-$ (m)			−2933	−2937	−2987	−3037
$t-665$ (s)			0.696	0.722	0.900	0.996
MLB reflected, steep						
$\widetilde{S}$ (s/km)	0.6442	0.6447		0.6473	0.6502	0.6510
θ_{MLB} (deg)	10.5	10.3		8.9	7.1	6.5
θ_{axial} (deg)	14.9	14.8		13.8	12.8	12.4
R (km)	52.43	52.22		51.06	50.00	49.75
T (s)	34.86	34.73		33.98	33.29	33.13
$\widetilde{z}^+$ (m)	−70	−70		−70	−70	−70
$\widetilde{z}^-$ (m)	−4244	−4191		−3881	−3544	−3443
$t-665$ (s)	0.058	0.203		0.563	0.864	0.983

VI. TIME FRONTS

An emitted acoustic pulse is received at a distant vertical hydrophone array as a series of discrete pulses, each represented by a "time front." The configuration of the time fronts is a function of the field of sound speed; perturbations in sound speed produce changes in the time front. The goal is to interpret the observed perturbations in the time fronts in terms of the ocean processes that cause them.

A. Refracted ray, no mixed layer

Figure 12 illustrates the properties of the canonical sound channel for an axial source and a vertical receiver array at 1000 km. Consider the time fronts +39 and +40, associated with 20 upper turning points (20 bounces) and with 19 and 20 LTPs, respectively, joined at the $+39\frac{1}{2}$ caustic.

The upper panels refer to the classical purely refracted case (no mixed layer). The left panel shows the first and last of the 20 complete ray loops, generated within a narrow beam of source angles extending from 11.1° to 9.6° (Table I). For ray **A**, the steepest, longest and deepest, the 20th loop

extends well beyond 1000 km, and has an UTP at the receiving array. The focus is on ray **C** with a LTP (and hence caustic) at the receiving array.

Time front +40 is generated by varying the source angle from 10.35° (point **C**) to 9.57° (point **E**), or the associated turning slowness from 0.6558 to 0.6574 s/km. For the axial point **D**,

$$r_0 = 1000 \text{ km} = NR(\widetilde{S}_0),$$

which yields

$$\widetilde{S}_0 = 0.6569 \text{ km}, \quad t_0 = NT(\widetilde{S}_0) = 665.873 \text{ s}. \tag{23}$$

For a chosen $\widetilde{S}$, the range and time for $N=20$ complete ray loops is

$$r = NR(\widetilde{S}), \quad t = NT(\widetilde{S}). \tag{24}$$

Suppose r exceeds r_0 by some fractional loop pR according to

$$r - r_0 = NR(\widetilde{S}) - NR(\widetilde{S}_0) = pR(S^*, \widetilde{S}), \tag{25}$$

where $S^*(z)$ is the slowness at the intersection of the ray with r_0. Equation (24) is solved for
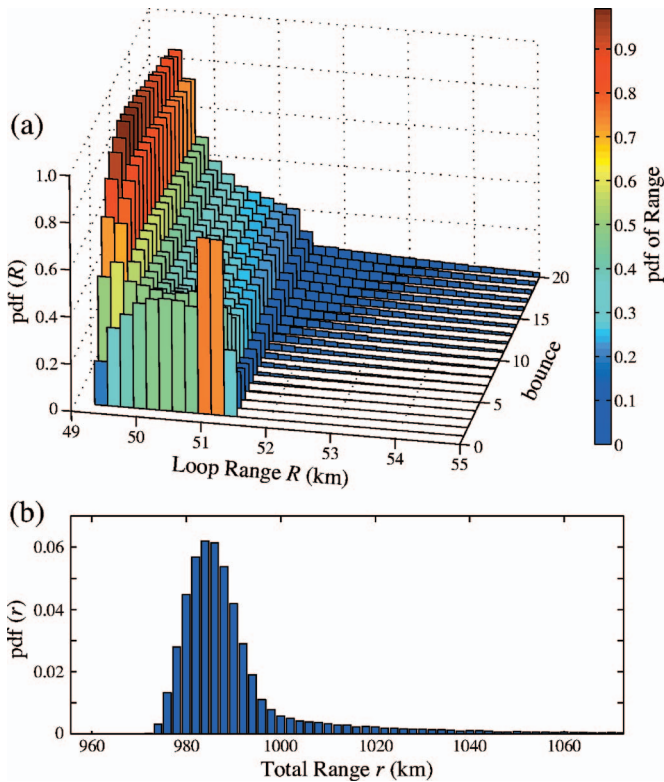
FIG. 10. (Color) (top) PDF of loop range R as a function of bounce for glancing incidence. (bottom) PDF of total range r from axis after 20 bounces to the lower turning point.

$$z^* = z(S^*), \quad S^* \equiv S^*(\widetilde{S}). \qquad (26)$$

Similarly

$$t - t_0 = NT(\widetilde{S}) - NT(\widetilde{S}_0) \qquad (27)$$

is the delay in axial crossing of ray $\widetilde{S}$ versus that of ray $\widetilde{S}_0$. Ray $\widetilde{S}$ crosses the receiver array at time t^*, with

$$t - t^* = pT(S^*, \widetilde{S}), \qquad (28)$$

so that
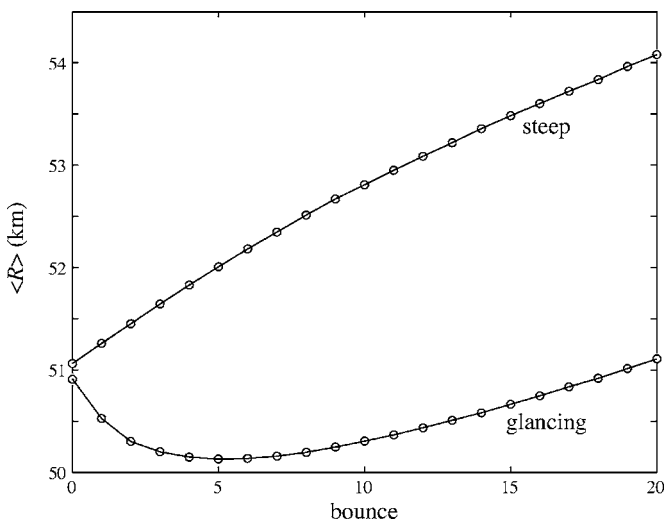
FIG. 11. Expected loop range as a function of bounce for glancing and steep incidence. Initial values are shown to the left.

$$t^* - t_0 = NT(\widetilde{S}) - NT(\widetilde{S}_0) - pT(\widetilde{S}). \qquad (29)$$

Thus each chosen $\widetilde{S}$ generates a point z^*, t^* of the time front in z,t space. An equivalent procedure with point **B** for reference yields time front +39. The solutions meet at the caustic ray **C** and are thus reconciled to yield the $39\frac{1}{2}$ V-pattern.

B. Mixed layer, plane MLB

The lower panels of Fig. 12 and Table I are for a mixed layer with a base taken (quite arbitrarily) at 70 m. There are now three pairs of time fronts with LTPs at 1000 km. Points **A**, **B** of the refracted time front lie above this depth. The refracted time front now starts with point **X**, corresponding to the ray tangent to the MLB. The left wing of the time fronts is accordingly clipped. Starting from the same point **X** is the reflected pair of time fronts **X**, **C′**, **D′**, **E′**. A third pair is designated **A″**, **B″**, **C″**, **D″**, **E″** in Fig. 12. All three caustics, **C**, **C′**, **C″** are at the receiver range r_0. The axial inclinations are 10.3°, 10.6°, and 13.8°; incident angles at the MLB are 0.6°, 8.9°, for the reflected rays, while the refracted ray turns below the MLB.

The refracted caustic is at a depth of 2.86 km; the glancing reflected caustic is only slightly deeper at 2.94 km, the steep reflected caustic is at 3.88 km. The refracted and glancing reflected caustics arrive within 3 ms, but the steep reflected arrives 0.16 s earlier. The steep reflected time front arrives earlier with increasing angle (Table I), as does the refracted time front. Interestingly, the glancing time front arrives later as angle increases, thus the progression through points in Fig. 12 is **X**, **C′**, **D′**, **E′**, **E″**, **D″**, **C″**, **B″**, **A″**.

The preceding discussion is for a mixed layer depth of 70 m, characteristic of early winter. Figure 13 shows the seasonal migration of the $+39\frac{1}{2}$ caustic. The trifurcation commences for a mixed layer of 7.2 m; for a MLB at 92 m the leading wing disappears (points **X** and **C′** in the previous figure overlap) and the caustic vanishes. The steep branch continues but the MLB becomes an increasingly weak reflector, and surface reflection (here neglected) takes over.

Thus the mixed layer formation is accompanied by a severe modification in the topology of the time fronts. The numbers cited are for purposes of discussion. The important consideration is that any seasonal variation of the depth of the lower caustic is of the same order, but quite different in origin, of the additional deepening associated with MLB waviness.

C. Mixed layer, wavy MLB

1. Single initial angle

Transmissions from launches at a single given incident angle, bouncing off a wavy MLB, generate an arrival pattern that is conveniently described by a joint PDF in time and depth. The arrival PDF for the glancing ray $\theta_0 = 0.0104$ is created using the Monte Carlo simulation of multiple bounces and summing individual arrivals after 20 bounces in bins 0.01 s in time and 0.05 km in depth [Fig. 14(a)]. For a flat MLB, the glancing ray arrives at point $C′$ on the range-independent time front. The wavy MLB causes most of the

D. L. Rudnick and W. H. Munk: Scattering from mixed layer into shadow
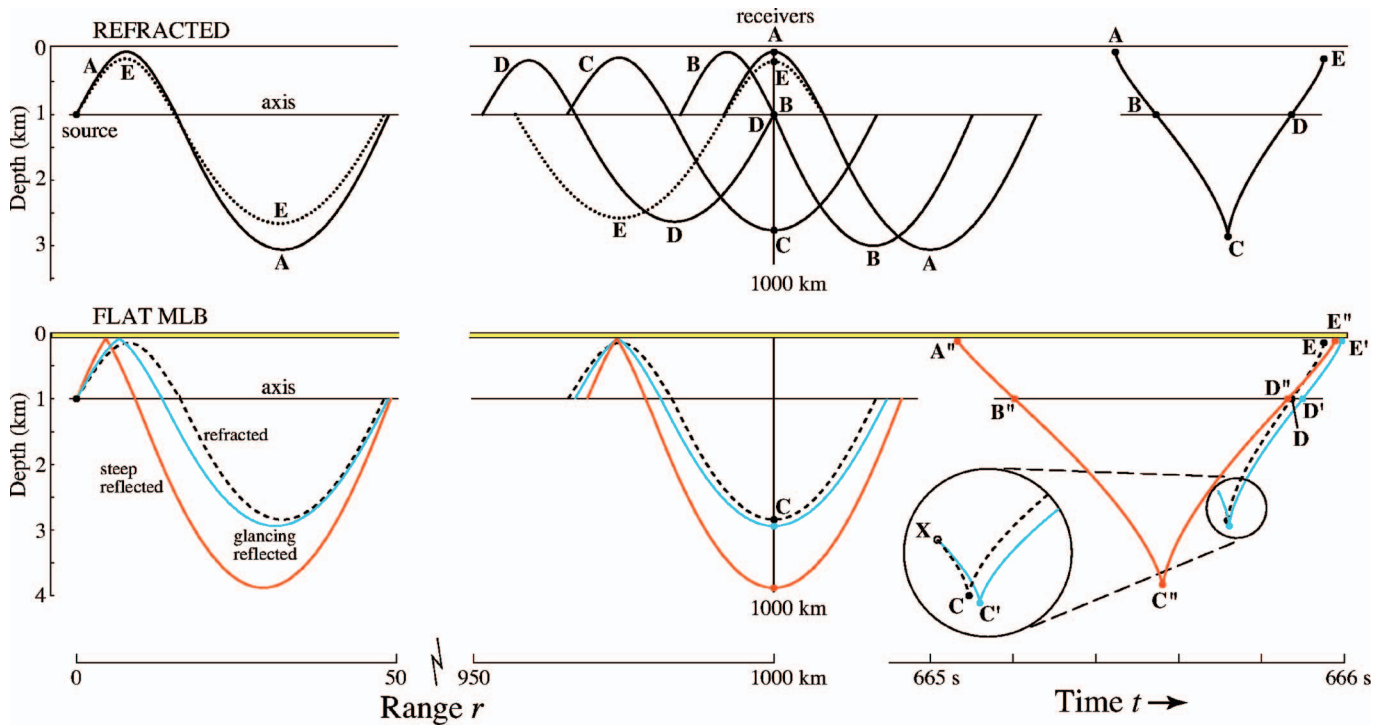
FIG. 12. (Color) Multiple time fronts from a mixed layer. The upper panels correspond to a canonical sound speed profile without mixed layer, the lower panels to a mixed layer of 70 m depth (thin yellow strip). **A** to **E** defines a narrow beam of launch angles; the first and last (20th) complete ray loops are shown, resulting in the time front to the right (black). In the presence of the mixed layer the left wing is clipped, resulting in a refracted time front **X** to **E** (black dashed). Two new pairs of time fronts (blue and red) are associated with reflections from the MLB.

arrivals to lay along the range-independent time front, just as in an experiment with a flat MLB and a range of initial angles. The spreading along the time front occurs across the
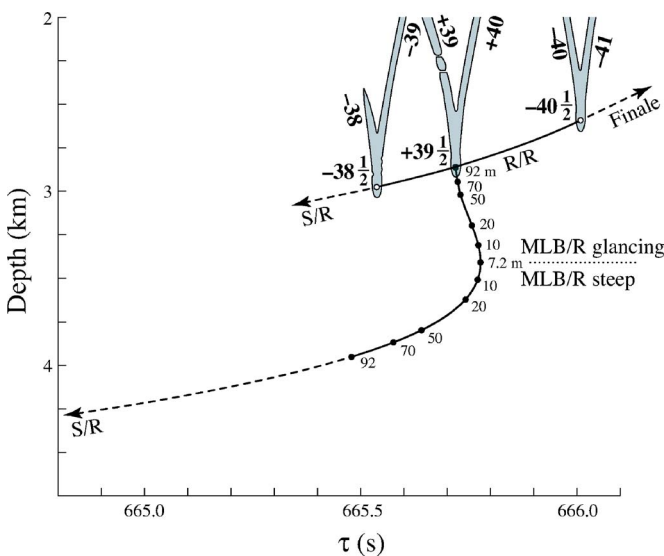
FIG. 13. Caustic migration for a 1000 km transmission through a canonical sound profile. R/R marks the location of the lower caustics ID $=-38\frac{1}{2}$, $+39\frac{1}{2}, -40\frac{1}{2}$ for (upper and lower) *refracted* turning points; they are preceded by surface *reflected* (S/R) rays (to the left, not shown) and followed by the axial finale (far to the right, not shown). The remaining plot deals with the $39\frac{1}{2}$ caustic (solid points). MLB/R marks the lower caustics of $+39\frac{1}{2}$ rays reflected at a MLB at stated depths. These lead to a triplication of $+39\frac{1}{2}$ arrivals starting with MLB at 7 m and ending with MLB$=92$ m (the upper turning point of the $+39\frac{1}{2}$ RR rays). The mixed-layer formation is associated with an overall deepening of the lower caustic from 2.9 km (R/R) to 3.9 km (MLB/R), and decrease in travel time by 0.3 s.

folds in the front caused by reflection off the MLB and refraction at depth. A wavy MLB also causes the region between the time fronts to be filled with arrivals, but spreading in the direction *across* the time front is much weaker than that *along* the time front. This asymmetry in spreading is a fundamental result of the model and related to the remarkable robustness of rays in a scattering environment. Arrivals appear at and below the lower turning point C'' for the steep ray. These arrivals well beneath the range-independent time front have similarities to the observations of Fig. 1. Also worth noting is the relative lack of arrivals along the extension C' to X of the time front to zero incident angle. Apparently these arrivals are not likely to be created by multiple bounces off the MLB. Out of the 100 000 transmissions in the Monte Carlo simulation, 95% are received at 1000-km range after exactly 20 bounces off the MLB. The remaining arrivals have a number of MLB bounces different from 20 and are not included in the PDF; most of them have anomalously long ranges (less than 20 bounces) and especially deep lower turning points.

The arrival PDF for the steep ray ($\theta_0 = 0.1555$) is calculated as for the glancing ray [Fig. 14(b)]. The steep ray arrives at C'' for a flat MLB. The time front is filled from E' to D' but not to the lower turning point C' of the glancing ray. The time front has a strong mode between E'' and C'', but some arrivals occur well beneath C''. The scattered arrivals occur *earlier* than for the flat MLB. In each case, the deep arrivals are associated with the tails in the distributions of $\cos \theta_r$. This pair of simulations presents the interesting asymmetry that the initial glancing ray produces arrivals at the lower turning point C'' of the steep ray, but the initial steep
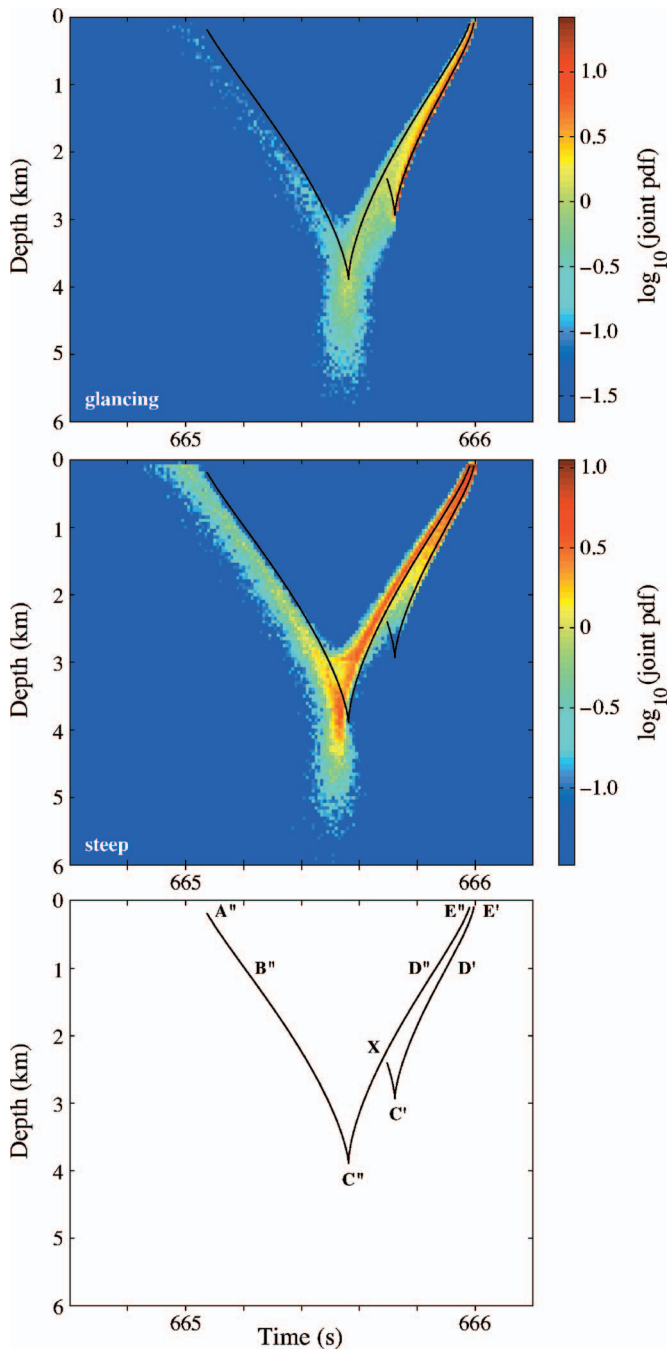
FIG. 14. (Color) Joint PDF of arrivals as a function of time and depth. PDFs are calculated from transmissions starting at glancing incidence (top) and steep incidence (middle). The black line is the range independent time front, with letters indicating certain points (as in Fig. 12) in the bottom panel. Note the penetration of arrivals below the range independent caustic.

FIG. 15. (Color) Joint PDF of arrivals as a function of time and depth. The range of initial incident angle is wide enough to include all rays that arrive on the shown branch of the time front. Three simulations are displayed: 1000-km range (top), 500-km range (middle), and 1000-km range with weak MLB slope variance of one-quarter the observed (bottom). Compare across-front scatter and penetration below the lower turning point in the three simulations.

ray does not produce arrivals at the lower turning point C' of the glancing ray. Most (58%) of the 100 000 transmissions are included in the PDF as they have exactly 20 bounces off the MLB, with the vast majority (92%) of the remainder having fewer bounces.

The simulations with fixed initial angles may be considered "stochastic Green's functions." For the flat MLB, arrivals are at single points (C' and C'', respectively); scattering from the wavy MLB spreads these points along defined time fronts. There is no explicit notion of ray intensity or attenuation in this very simple model. So the PDF is properly in-

terpreted as a measure of the likelihood of a reception, although the PDF bears a resemblance to observations of intensity. This resemblance suggests that observed intensity is a strong function of the density of arrivals.

D. Wide range of initial angles

Simulations using a wide range of initial angles are most comparable to observations using a nondirectional source (Fig. 15). Three simulations are done at 1000- and 500-km

range for $\sigma^2 = 2 \times 10^{-4}$, and at 1000-km range for $\sigma^2 = \frac{1}{2} \times 10^{-4}$. A range of initial angles from 0 to 0.4, by steps of 0.001, are used, and at each step 500 realizations are done. The steepest initial angle is chosen to be large enough that all of the transmissions are past the receiver before the 20th (for 1000-km range) or 10th (for 500-km range) bounce, so none appear in the PDF; using steeper initial angles up to $\pi/2$ would add nothing to the following results. In the interest of simplicity, the model sound channel is taken to be infinitely deep, so bottom effects are not considered. Whether a transmission is included in the arrival PDF is determined only by the number of bounces off the MLB it has experienced, not whether it has encountered the bottom.

For the 1000-km range, the arrival pattern is something of a superposition of the patterns from the glancing and steep rays [Fig. 15(a)]. Features in this all-angle PDF are similar to those for the glancing and steep rays, suggesting again the influence of along-time-front scattering. If the effect of a wavy MLB is important in the ocean, an arrival pattern similar to this one should be observed.

At 500-km range, the most obvious difference is the reduction in across-time-front scatter [Fig. 15(b)]. Penetration of arrivals beneath the lower turning point of the range-independent ray is only marginally deeper at the 1000-km than at 500-km range, suggesting that penetration is a relatively weak function of range. The general features of the 1000- and 500-km simulations may be compared to observations to address the plausibility of the scattering processes in the simulation.

The final simulation is at the 1000-km range, but with weak MLB slope variance of one-quarter the observed value. Across-time-front scatter and penetration beneath the lower turning point are both reduced compared to the full-strength simulation. The implication is that penetration into the shadow zone is a strong function of MLB slope variance.

E. Robustness in a scattering environment

In the early days of ocean acoustic tomography, the discovery that time fronts maintained their integrity in a scattering environment came as a pleasant surprise (Spiesberger *et al.* 1980; Munk *et al.*, 1995, p. 362). It should not have been; it is a manifestation of the principle discovered by Fermat in 1655 that between fixed source and receiver a ray path is defined by a parabolic extremum in travel time, so that travel time does not vary to first order with perturbation of the path.

The time fronts in Fig. 14 were constructed by two quite distinct methods: the color coding designates the density of "hits" at the 1000-km range in 100 000 Monte Carlo realization of the wavy MLB, using one (or many) fixed launch angles. The black lines were generated for a range-independent ocean by varying the launch angles and deriving the time and depth at the 1000-km range in accordance with Eqs. (24) to (29). There is a striking resemblance between the two time fronts so derived (the small but systematic displacement is of course crucial).

There have been extensive discussions of the stability of the early arrivals. An important variable is the second action
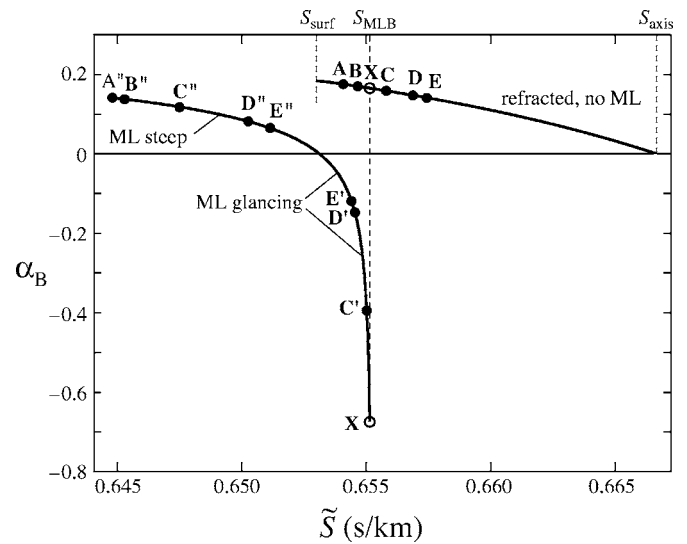
FIG. 16. The stability parameter $\alpha_B(\widetilde{S})$. For the canonical profile (with no mixed layer) the upper stability curve continues to the left until surface reflection, with the points **A** to **E** corresponding to the points shown in Fig. 12 on the pair of time fronts +39 and +40 emerging from the lower caustic $+39\frac{1}{2}$. Point **X** has an upper turning point at $z = -70$ m. For a surface mixed layer of 70 m thickness, the $+39\frac{1}{2}$ pair of refracted time fronts is shortened to **X** to **E**. Two new pairs of $+39\frac{1}{2}$ time-fronts lie along a second branch of $\alpha_B(\widetilde{S})$: **X**, **C′**, **D′**, **E′** for near-grazing incidence on the MLB, and **A″** to **E″** for steep incidence.

derivative A'' (for positive A'' the scattered signal arrives early). Brown *et al.* (2005) discuss the ray-based "stability parameter" α_B and the mode-based "waveguide invariant" β_B:

$$\alpha_B(\widetilde{S}) = \frac{A(\widetilde{S})A''(\widetilde{S})}{A'^2(\widetilde{S})}, \quad \beta_B = \frac{ds_{\text{gr}}}{ds_{\text{ph}}} \tag{30}$$

and show them to be equivalent in asymptotic mode theory[2] (adding the subscript B to distinguish these from the variables introduced earlier). This parameter plays a central role in long-range propagation through a range-independent environment with weak range-dependent scatterers (see a recent volume edited by Kuperman and D'Spain 2002). The group slowness s_{gr} is proportional to t, and the phase slowness $s_{\text{ph}} = \widetilde{S}$ is closely related to the vertical source angle, so $\beta \sim dt/d\theta \rightarrow 0$ indicates a focusing in time of vertical aperture. Figure 16 shows the stability parameter for a canonical ocean with a reflecting MLB at 70 m. The three branches correspond to the trifurcation. The zero at the termination of α_B is related to the crescendo in the SOFAR finale. The zero crossing of the reflection branch may indicate a prominent observable. (Cusps are related to $dt/dz \rightarrow 0$.)

The stability parameter helps one to understand some of the features of the arrival PDFs (Figs. 14 and 15). The density of arrivals should increase with decreasing $|\alpha_B|$. The lack of arrivals near the beginning of the time front (points **X**, **C′**) is thus related to the large value of $|\alpha_B|$ there, and the high density of arrivals near the MLB (points **E′**, **E″**) is a region of small $|\alpha_B|$. Positive α_B implies anomalously early arrivals

as seen in the left branch of the steep time front (points **A″**, **B″**, **C″**). A more complete investigation of α_B is warranted, but is beyond the ability of the authors.

VII. DISCUSSION

The incentive for this study was provided by recent profiling of the upper ocean. Three sections in the North East Pacific further emphasized the sharp transition at the base of the wind-mixed layer. The layer thickness varies greatly between winter and summer, and its base is distorted by internal waves.

The properties of the MLB play a crucial role in long-range acoustic propagation; in particular, its waviness is associated with a random walk toward deepening turning points of rays in the ocean sound channel. This may account for the long-known experience that acoustic signals are received at depths well exceeding those expected to be insonified (Fig. 1). Screening (incident rays insonifying only part of a wavy boundary) tends to steepen rays. However screening is not essential to the deepening of scattered rays, which is determined by variance in MLB slope and the resulting random walk in reflected angle. The increase in $\langle \cos \theta_r \rangle$ is nearly the same for both glancing and steep rays, whereas screening is negligible for initial angles much greater than the root-mean-square MLB slope (Fig. 9).

Two statistics characterize the deepening in fundamentally different ways. The random walk in reflected angle causes a mean deepening per bounce of the lower turning point by 35 m using observed MLB slope variance and abyssal adiabatic sound speed gradient. So after 20 bounces off the MLB the mean deepening is 700 m. The average that produces this result is over all rays, regardless of their total range. A distinct statistic is the deepening at the lower turning point for given total range (Figs. 14 and 15). In contrast to the deepening for a given number of bounces, the deepening at a given range is a weak function of range and a strong function of MLB slope variance. An analytical treatment of the deepening at given range has not yet been achieved, however the Monte Carlo results seem solid.

Many simplifying assumptions have been made: a ray-optical formulation with a reflecting MLB, a high-wave-number cutoff of the internal wave spectrum, some properties of acoustic screening by the steep internal wave slopes, etc. The shadow penetration of later arrivals with upper turning points below the MLB calls for a more general random walk process, and is presumably associated with the waviness of iso-velocity contours near the upper turning points.

Surface reflections determine the very start of an incoming SOFAR gram, followed by features controlled by mixed layer parameters. Both boundaries, surface and MLB, lead to a trifurcation of the incoming ray-fronts (Dzieciuch *et al.*, 2004). Together they determine the complex structure; the SOFAR overture remains to be composed.

Understanding the penetration into the wave shadow in terms of a simple model provided the focus for this paper. But the motivation is not only for solving an old puzzle. We anticipate monitoring upper ocean processes by abyssal acoustic arrays. This includes not only the profiles of sound speed (temperature) and velocity by the method of ocean acoustic tomography, but also the waxing and waning of the mixed layer, frontal formation, and internal wave intensity.

The last few decades have seen remarkable progress by remote sensing of the ocean from above. But there is much to be learned by "viewing" the ocean from below.

APPENDIX A: TRANSFORMATION OF GM SPECTRUM

The mean-square of vertical displacement is written (Munk, 1981)

$$\langle \zeta^2 \rangle = \sum_j \int_0^{-\infty} d\omega \, F_\zeta(\omega, j),$$

$$F_\zeta(\omega, j) = b^2 \frac{N_0}{N} \frac{\omega^2 - f^2}{\omega^2} E(\omega, j),$$

$$E(\omega, j) = 2\pi^{-1} f \omega^{-1} (\omega^2 - f^2)^{-1/2} (j^2 + J^{*2})^{-1} JE, \quad (A1)$$

where $b = 1300$ m, $N_0 = 3$ cph, $f = \frac{1}{24}$ cph, $E = 6.3 \times 10^{-5}$, and $J = 2.1$ (Garrett and Munk, 1972a). Performing the summation and integration yields

$$\langle \zeta^2 \rangle = \tfrac{1}{2} b^2 E (N/N_0)^{-1} = 53 (N/N_0) \text{m}^2 \quad (A2)$$

or 7.3 m rms near the surface.

Going from vertical mode j to horizontal wave number κ:

$$\kappa = \sqrt{\kappa_1^2 + \kappa_2^2} = \pi b^{-1} \left(\frac{\omega^2 - f^2}{N_0^2 - \omega^2} \right)^{1/2} j \quad (A3)$$

yields

$$F_\zeta(\omega, \kappa) = F_\zeta(\omega, j) dj/d\kappa = \frac{2b^3 E}{\pi^2} f (N_0^2 - \omega^2)^{1/2} \omega^{-3} \frac{J}{j^2 + j^{*2}}. \quad (A4)$$

Substituting for j from Eq. (A2), assuming $j \gg j^* = 3$,

$$F_\zeta(\omega, \kappa) = 2bEJI(\omega)\kappa^{-2}, \quad I(\omega) = \frac{f(\omega^2 - f^2)}{(N_0^2 - \omega^2)^{1/2} \omega^3}. \quad (A5)$$

Integrating over frequency,

$$F_\zeta(\kappa) = \int_f^{N_0} F_\zeta(\omega, \kappa) d\omega = 2bEJ\Omega\kappa^{-2}, \quad (A6)$$

where

$$\Omega \equiv \int_f^{N_0} I(\omega) d\omega = \frac{f}{2N_0} \left[2 \log \frac{2N_0}{f} - 1 \right]$$

for $f \ll N_0$. $\Omega = 0.062$ for $f/N_0 = 1/72$. Assuming horizontal isotropy

$$F_\zeta(\kappa) d\kappa = F_\zeta(\kappa_1, \kappa_2) d\kappa_1 \, d\kappa_2 = F_\zeta(\kappa_1, \kappa_2) \kappa \, d\kappa \, d\theta$$
$$= 2\pi F_\zeta(\kappa_1, \kappa_2) \kappa \, d\kappa,$$

$$F_\zeta(\kappa_1, \kappa_2) = \frac{1}{2\pi\kappa} F_\zeta(\kappa) = \pi^{-1} bEJ\Omega\kappa^{-3}. \tag{A7}$$

Note that $\int_{-\infty}^{\infty} d\kappa_2 \kappa^{-3} = \int_{-\infty}^{\infty} d\kappa_2 (\kappa_1^2 + \kappa_2^2)^{-3/2} = 2\kappa_1^{-2}$ yields the *towed* spectrum

$$F'_\zeta(\kappa_1) = \int_{-\infty}^{\infty} d\kappa_2\, F_\zeta(\kappa_1, \kappa_2) = 2\pi^{-1} bEJ\Omega\kappa_1^{-2} \tag{A8}$$

for $-\infty < \kappa_1 < +\infty$. The *one-dimensional* spectrum is $F_\zeta(\kappa_1) = F'_\zeta(\kappa_1) + F'_\zeta(-\kappa_1) = 2F'_\zeta(\kappa_1)$ with $0 < \kappa_1 < \infty$. Then

$$\langle\zeta^2\rangle = \int_{\kappa_1^*}^{\infty} d\kappa\, F_\zeta(\kappa_1), \quad F_\zeta(\kappa_1) = K(\kappa_1^*)^{-2},$$

$$K = 4\pi^{-1} bEJ\Omega = 0.0136 \text{ m},$$

$$\kappa_1^* = 8\pi^{-1} J\Omega b^{-1} = 2.55 \times 10^{-4} \text{ m}^{-1}, \tag{A9}$$

where the cut-off κ_1^* has been so chosen that $\langle\zeta^2\rangle = K/\kappa_1^* = \frac{1}{2}b^2E = 53 \text{ m}^2$ in agreement with Eq. (A2).

The associated one-dimensional slope spectrum is

$$F_m(\kappa_1) = \kappa_1^2 F_\zeta(\kappa_1) = 4\pi^{-1} bEJ\Omega \text{ rad}^2/\text{rpm}. \tag{A10}$$

Finally going from radians to cycles per meters, $\kappa_1 = 2\pi\tilde{\kappa}_1$,

$$F_m(\tilde{\kappa}_1) = \frac{d\kappa_1}{d\tilde{\kappa}_1} F_m(\kappa_1) = 8bEJ\Omega = 0.086 \text{ rad}^2/\text{cpm}, \tag{A11}$$

compared to 0.02 rad^2/cpm measured.

APPENDIX B: THE ACTION VARIABLE $A(\tilde{S})$

The action variable is defined by

$$A(\tilde{S}) = 2\int_{\tilde{z}^-}^{\tilde{z}^+} dz\sqrt{S^2(z) - \tilde{S}^2}. \tag{B1}$$

For a given sound profile $S(z)$, A is a function of the ray variable $\tilde{S}$. Write A' for $dA/d\tilde{S}$, etc. It can be shown that (Munk *et al.*, 1995)

$$R = -A', \quad R' = -A'',$$

$$T = A + R\tilde{S}, \quad T' = -\tilde{S}A'', \quad T'' \approx -A''. \tag{B2}$$

Consider *partial* loops, from their axial crossing to some depth z between axis and turning points. These can be derived from the partial action

$$pA(S, \tilde{S}) = \int_{z_{ax}}^{z} dz\sqrt{S^2(z) - \tilde{S}^2} \tag{B3}$$

according to

$$pR = -(pA)', \quad pT = pA + \tilde{S}pR. \tag{B4}$$

APPENDIX C: PDF FOR MULTIPLE BOUNCES

It is possible to calculate the joint PDF of inclinations of the MLB from multiple bounces. First consider two bounces. The PDF of MLB inclination α_1 for the first bounce is

$$f(\alpha_1) = \begin{cases} \dfrac{p(\alpha_1)}{D(\theta_0)}, & \alpha_1 < \theta_0 \\[2mm] 0, & \alpha_1 \geq \theta_0, \end{cases} \tag{C1}$$

where θ_0 is the inclination of the incident ray. Note that θ_0 is not a random variable. The conditional PDF of the MLB inclination α_2 for the second bounce given the first MLB inclination α_1 is

$$f(\alpha_2|\alpha_1) = \begin{cases} \dfrac{p(\alpha_2)}{D(\theta_0 - 2\alpha_1)}, & \alpha_2 < \theta_0 - 2\alpha_1 \\[2mm] 0, & \alpha_2 \geq \theta_0 - 2\alpha_1, \end{cases} \tag{C2}$$

where the fact has been used that the first reflection will arrive for the second bounce with an incident angle of $\theta_0 - 2\alpha_1$. Given these distributions, the joint distribution of α_1 and α_2 is (Papoulis, 1991)

$$f(\alpha_1, \alpha_2) = f(\alpha_2|\alpha_1)f(\alpha_1). \tag{C3}$$

The resulting PDF is a normalized joint Gaussian with sharp cutoffs defined by $\alpha_1 = \theta_0$ and $\alpha_2 = \theta_0 - 2\alpha_1$.

The joint PDF of the α's for N bounces can be written as

$$f(\alpha_1, \alpha_2, \ldots, \alpha_N)$$
$$= f(\alpha_N|\alpha_{N-1}, \alpha_{N-2}, \ldots, \alpha_1) \ldots f(\alpha_2|\alpha_1)f(\alpha_1) \tag{C4}$$

where, for $N \geq M \geq 2$,

$$f(\alpha_M|\alpha_{M-1}, \alpha_{M-2}, \ldots, \alpha_1)$$
$$= \begin{cases} \dfrac{p(\alpha_M)}{D\left(\theta_0 - 2\sum\limits_{j=1}^{M-1}\alpha_j\right)}, & \alpha_M < \theta_0 - 2\sum\limits_{j=1}^{M-1}\alpha_j \\[5mm] 0, & \alpha_M \geq \theta_0 - 2\sum\limits_{j=1}^{M-1}\alpha_j. \end{cases} \tag{C5}$$

Of interest is the PDF of the sum of the α's,

$$\gamma = \sum_{j=1}^{N} \alpha_j, \tag{C6}$$

which can, in principle, be calculated by making a substitution like

$$\alpha_N = \gamma - \sum_{j=1}^{N-1} \alpha_j \tag{C7}$$

into Eq. (C4) and then integrating off all the α_j's. Because γ is a sum of random variables, its PDF will approach a normal distribution as N approaches infinity. But the integration of Eq. (C4) does not appear to be analytically tractable.

---

[1]Here the exact relation (Appendices (B1)–(B4)) is used rather than the perturbation expansion (21). Referring to Munk *et al.* (1995) 44-47 and 106-114, in Eq. (2.18.24) replace $\mp$ by $\pm$, and do *not* set $\tilde{S}/S_A = 1$. Action for a reflecting upper boundary is derived in Dzieciuch *et al.* (2004) appendix.

[2]Brekhovskikh and Lysanov (1991) refer to A as the "ray invariant."

Beckmann, P., and Spizzichino, A. (**1987**). *The Scattering of Electromag-

netic Waves from Rough Surfaces (Artech House, Boston).

Brekhovskikh, L., and Lysanov, Y. (**1991**). *Fundamentals of Ocean Acoustics* (Springer, Berlin).

Brown, M., Beron-Vera, F., and Rypina, I. (**2005**). "Rays, modes, wavefield structure, and wavefield stability," J. Acoust. Soc. Am. **117**, 1607–1610.

Dzieciuch, M. (private communication).

Dushaw, B. D., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1999**). "Multimegameter-range acoustic data obtained by bottom-mounted hydrophone arrays for measurement of ocean temperature," IEEE J. Ocean. Eng. **24**(2), 202–214.

Dzieciuch, M., Munk, W., and Rudnick, D. L. (**2004**). "The propagation of sound through a spicy ocean; The SOFAR overture," J. Acoust. Soc. Am. **116**(3), 1447–1462.

Flatté, S., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, New York).

Garrett, C., and Munk, W. (**1972a**). "Space-time scales of internal waves," Geophys. Fluid Dyn. **2**, 225–264.

Hodges, B. A., and Rudnick, D. L. (2006). "Horizontal variability in chlorophyll fluorescence and potential temperature," Deep-Sea Res., Part I (in press).

Klymak, J. M., and Moum, J. N. (2006). "Interpreting horizontal temperature spectra in the ocean," J. Phys. Oceanogr. (in press).

Kuperman, W., and D'Spain, G. (editors) (**2002**). Ocean Acoustic Interference Phenomena and Signal Processing, *American Institute of Physics Conference Proceedings* Vol. **621** (AIP Press, New York).

Munk, W. H. (**1981**). "Internal waves and small scale processes," in *Evolution of Physical Oceanography-Scientific Surveys in Honor of Henry Stommel*, edited by B. Warren and C. Wunsch (MIT, Boston), pp. 264–291.

Munk, W., Worcester, P., and Wunsch, C. (**1995**). *Ocean Acoustic Tomography* (Cambridge University Press, New York).

Papoulis, A. (**1991**). Probability, Random Variables, and Stochastic Processes, *Communications and Signal Processing*, 3rd ed. (McGraw-Hill, New York), 666 pp.

Phillips, O. M. (**1966**). *The Dynamics of the Upper Ocean* (Cambridge University Press, New York).

Rudnick, D. L., Klinke, J., and Hodges, B. A. (**2004**). "The Underway CTD," EOS Trans. Am. Geophys. Union **85**(47), Fall Meet. Suppl., Abstract OS21C-1244.

Spiesberger, J. L., and Tappert, F. D. (**1996**). "Kaneohe acoustic thermometer further validated with rays over 3700 km and the demise of the idea of axially trapped energy," J. Acoust. Soc. Am. **99**, 173–184.

Spiesberger, J. L., Spindel, R. C., and Metzger, K. (**1980**). "Stability and identification of ocean acoustic multipaths," J. Acoust. Soc. Am. **67**, 2011–2017.

Worcester, P. (**1977**). "Reciprocal acoustic transmissions in a midocean environment," J. Acoust. Soc. Am. **62**, 895–905.

Measurements of temporal coherence of sound transmissions through shallow water

T. C. Yang[a)]

Naval Research Laboratory, Washington, DC 20375

In this paper we report the measurements of temporal coherence of acoustic signals propagating through shallow water using data from three experiments in three different parts of the world, with sound speed standard deviation (STD) varying from 0.3 to 5 m/s near the layer depth. Temporal coherence is estimated from the autocorrelations of broadband channel impulse functions, the latter are deduced from broadband signals transmitted through the ocean during the experiments. The measurements covered three frequency bands: low frequencies below 1.2 kHz, midfrequencies between 2 and 5 kHz, and high frequencies between 18 and 22 kHz. The source-receiver range covers 3, 5, 10, and 42 km. The signal coherence-time is defined and deduced from the data. Motivated by previous theoretical work in deep water on the signal coherence-time as a function of the signal frequency, the source-receiver range, and sound speed STD, a similar but empirical analysis is applied to the measured data in shallow water. While the range dependence agrees with the theory, the data exhibit a different dependence on the signal frequency than the theoretical prediction for deep water. [DOI: 10.1121/1.2345910]

I. INTRODUCTION

The ocean is a dynamic (time varying/random) medium. As a result, acoustic signals propagating through the ocean fluctuate with time. The rate of the signal fluctuation is often described by the signal temporal coherence (i.e., the autocorrelation of the signal normalized by the average signal energy) that measures in a statistical sense the rate of decrease of the signal coherence. The higher the signal fluctuation, the faster the temporal coherence decreases with time. The signal coherence-time is defined as the time by which the signal becomes "uncorrelated" (as when the temporal coherence drops to $1/e$).

Knowing the signal coherence-time is useful for many practical applications. For example, for the detection and localization of narrow band acoustic signals, the signal coherence-time determines the maximum time the signal can be integrated (before losing coherence), which, in turn, determines the maximum temporal processing gain. For matched field processing, the autocorrelation time between the initial field (used as the replica field) and the data field determines for how long the replica field can be used. For adaptive (temporal) signal processing, such as an adaptive channel equalizer in underwater acoustic communications, the rate of signal adaptation is controlled by the temporal stability of the signal or the signal coherence-time.

The practical need has prompted many experimental and theoretical studies of the temporal coherence of acoustic signals. Many experimental works reported in the literature were concerned with low frequency signals propagated to long ranges in deep oceans. Spindel *et al.*[1] and Clark *et al.*[2] focused on signal propagation at 406 Hz up to 250 km.

Williams[3] and Williams *et al.*[4] studied signal propagation at 400 Hz up to 1200 km. Georges *et al.*[5] and Birdsall *et al.*[6] investigated 57 Hz signal propagation to thousands of kilometers. Fitzgerald *et al.*[7] focused on very low frequency (10 Hz) signals. Dyer *et al.*[8] studied a band of frequencies (20–200 Hz) in the Marginal Ice Zone. The measurement results from these experiments can be expressed in terms of the signal phase rate, which is the inverse of the signal coherence-time, as discussed later. Temporal coherence was modeled theoretically using theories of signal propagations in a random medium. Earlier work calculated the signal phase rate using the Rytov approximation.[9] The emphasis was on the influence of internal waves in the deep ocean using the Garrett-Munk internal wave model.[10] The model predicted that the signal coherence-time is proportional to the inverse first power of frequency, inverse square-root power of range, and inversely proportional to the amplitude of the internal waves or the sound speed standard deviation (STD).

In contrast to the abundant work in deep water, there were few measurements on the temporal coherence of acoustic signals propagating in shallow water until recently. Using the ADVENT99 data collected in the Strait of Sicily, Nielsen *et al.* reported snapshot measurements of temporal coherence using linear frequency modulated (LFM) signals.[11] From the same data, using multitone signals, Yang *et al.* reported the temporal coherence of the acoustic signals using matched-field correlation of data fields.[12,13] By decomposing the data into mode space, they also measured the temporal coherence of individual (low) order modes. Both the LFM and multitone showed a signal coherence-time on the order of 20–200 min at frequencies below 1 kHz. Headrick *et al.* deduced the mode correlation from SWARM 95 data received on the WHOI vertical array at a range of 33 km.[14] Rouseff *et al.* measured and modeled the modal temporal coherence us-

[a)]Electronic mail: yang@wave.nrl.navy.mil

ing the data received on the NRL array at a range of 42 km.[15] Mignerey *et al.* reported a significant variation of the matched-field correlation of low frequency signals in the South China Sea, depending very much on the presence or absence of the solitary internal waves.[16] Badiey *et al.* measured the temporal coherence of acoustic signals in very shallow (~15 m) regions.[17]

Given the recent interest on signal propagation in shallow water, and the broadening of interest from low to high frequencies (such as for acoustic communications), the need exists for a systematic measurement of the temporal coherence of signals propagated through shallow waters to determine its relationship to the oceanographic environment. We report in this paper experimental measurements of temporal coherence in three different shallow water areas:[18] Sicily Strait in the Mediterranean, coastal water off New England, and coastal water off New Jersey, covering a frequency range from 200 Hz to 22 kHz and a propagation range from 3–4 km to 40 km. Issues critical to the signal coherence measurement are addressed.

From the measured signal coherence, one can estimate the signal coherence-time as defined later. A question of interest is what is the frequency and range dependence of the signal coherence-time in shallow water (as compared with deep water). For this study, an empirical analysis is used, noting the lack of theoretical predictions for shallow water. Another question of interest is what causes the signal coherence loss. In general, signal coherence loss arrives from sound scattering from the moving sea surface and time-varying oceanographic processes. (Signal coherence-time is infinite in a frozen ocean.) For a bottom mounted source in a downward refractive sound speed profile, acoustic propagation modeling results suggest that the contribution of the surface is minimal (see later). Assuming medium fluctuation is the dominant cause for the coherence loss, an effort is made to determine the relation between the signal coherence-time with the sound-speed STD, using conductivity-temperature-depth (CTD) data taken during the acoustic experiment.

To properly interpret the coherence-time results, one needs to model the influence of the various oceanographic processes (internal waves, turbulence, etc.) on the acoustic signal. Such a study is beyond the scope of this paper. Note that if the signal fluctuation can be described by one statistical distribution (as when there exists only one oceanographic process), the temporal coherence measurement will be independent of time and the size of the ensemble as long as the latter is statistically significant. In the real ocean, there exist various oceanographic processes that evolve with different time scales.

Consequently, the signal coherence measured at one time could be different than that measured at another time. Thus, in addition to the average temporal coherence of the signal, one needs to also study the "instantaneous" temporal coherence of the signal—the former is the average over samples covering a long period of time (hours/days), and the latter is the average over samples covering a short period of time (seconds/minutes). The "instantaneous" signal coherence-time provides a measure (variance) of the tempo-

ral variation of the signal coherence-time; the variance is a function of the acoustic frequency. The mean and variance of the "instantaneous" signal coherence-time can be used to infer the properties of the oceanographic processes responsible for the signal coherence loss.

The data analyzed involve acoustic signals of different carrier (center) frequencies with different bandwidths. Naturally, to compare the results, one needs to ask whether the measurements are influenced by the signal bandwidth. Note that the signal coherence for a (extra) wide band signal is not likely to be the same as that of a narrow band signal. When the signal bandwidth is small compared with the signal frequency, most works have assumed that the signal bandwidth is not an issue with which to be concerned. The bandwidth dependence is studied below by estimating the temporal coherence with a bandpass filter (with different bandwidths) applied to data with a 2 kHz bandwidth. The results showed little bandwidth dependence.

This paper is organized as follows. In Sec. II we describe the definitions and measurement techniques. In Sec. III we illustrate the measurements of the signal temporal coherence at three frequencies and two ranges using the AS-COT01 data, including a discussion on the measurement issues. In Sec. IV we present the results of signal temporal coherence measured from the SWARM95 experiment. In Sec. V we present the results of signal temporal coherence measured from RAGS03 experiment. The range and frequency dependence of the signal coherence-time are studied in Sec. VI. Environmental impacts on the signal temporal coherence are briefly discussed. Section VII is a short summary.

II. DEFINITIONS AND MEASUREMENT TECHNIQUES

Temporal coherence of a signal is defined by the (auto-) correlation of the signals separated by a delay time τ, normalized by the power of the signal, as given by

$$\rho(\tau) \equiv \frac{\langle p^*(t)p(t+\tau)\rangle}{\sqrt{\langle p^*(t)p(t)\rangle\langle p^*(t+\tau)p(t+\tau)\rangle}}, \tag{1}$$

where $\langle \rangle$ denotes the ensemble average and * denotes the complex conjugate. The ensemble average is usually done over time t using the ergodic theorem. The ensemble average can also be over frequencies (i.e., different narrow band signals), or space (i.e., sensors over an array), or combinations of all the above.

Earlier work on deep water data was focused on the phase rate of a narrow band signal.[1–8] A narrow band signal can be represented after quadrature demodulation (i.e., after removing the carrier frequency) as

$$p(t) = R(t)e^{i\phi(t)}, \tag{2}$$

where $R(t)$ is the amplitude and $\phi(t)$ is the phase. It was assumed that the signal variation is predominantly in its phase. Ignoring the amplitude variation, i.e., $R(t+\tau) \approx R(t)$, one finds that the autocorrelation of the pressure field can be approximated by[19,20]

$$\langle p^*(t)p(t+\tau)\rangle \simeq \mu^2\langle e^{-i\phi(t)}e^{i\phi(t+\tau)}\rangle \simeq \mu^2 e^{(-1/2)\dot\phi^2 r^2}, \qquad (3)$$

where τ is the delay time, $\dot\phi$ denotes the time derivative of the signal phase (phase rate), and μ^2 denotes the mean intensity $\langle R(t)R(t+\tau)\rangle \simeq \mu^2$. Using Eq. (3), the temporal coherence as defined by Eq. (1) becomes

$$\rho(\tau) = e^{(-1/2)\dot\phi^2 r^2}. \qquad (4)$$

While noting that the assumptions made above for deep water may not be valid for shallow water, one observes at the same time that the functional dependence of the signal coherence [Eq. (4)] has been widely assumed for data analysis. It will be used later and seems to be valid for a small delay time.

The signal coherence-time, commonly defined by the delay time at which $\rho(\tau)=1/e$, is given by

$$\tau_{1/e} = \sqrt{2}(\dot\phi)^{-1}. \qquad (5)$$

In terms of $\tau_{1/e}$, Eq. (4) becomes $\rho(\tau)=e^{-\tau^2/\tau_{1/e}^2}$. Defining $\dot\phi=2\pi\nu$, one finds that ν is the Doppler spread for a narrow band signal. This can be verified by taking a Fourier transform of the correlation function, Eq. (3). It is evident from Eq. (5) that the Doppler spread is inversely proportional to the signal coherence-time.

For reasons having to do with data analysis (as discussed later in Sec. III), one finds it necessary to measure the signal coherence-time while the signal coherence is high. A practical choice is to define $\tau_{0.8}$, which is the coherence-time at $\rho=0.8$. It is related to the Doppler spread by

$$\nu = \sqrt{\frac{-\ln(0.8)}{2\pi^2}}\,\tau_{0.8}^{-1} \simeq 0.11\tau_{0.8}^{-1}. \qquad (6)$$

$\tau_{0.8}$ has been previously discussed in the context of underwater acoustic communications[18] as communications often require a signaling period with high signal coherence.

For deep water, using the Rytov approximation, earlier theoretical models have suggested that ν should increase linearly with the acoustic frequency and square-root power of the source-receiver range,[9]

$$\nu = \gamma f\sqrt{R}, \qquad (7)$$

where the coefficient γ is proportional to the amplitude of the internal waves. This suggests that the coherence-time $\tau_{0.8}$ should decrease linearly with the inverse of the acoustic frequency and inverse square root of the range. This dependence will be tested with the shallow water data analyzed later.

A second method to measure the temporal coherence of a (narrow band) signal uses matched field correlation.[12,16] A (narrow band) signal field at time t is used as the reference (replica) field and correlated with the signal that follows at a later time, $t+\tau$. It is called matched field correlation since the ensemble average is over the sensors on an array,

$$\rho(t,\tau) = \frac{\displaystyle\sum_{j=1}^{N} p_j^*(t)p_j(t+\tau)}{\sqrt{\displaystyle\sum_{j=1}^{N} p_j^*(t)p_j(t)\sum_{j=1}^{N} p_j^*(t+\tau)p_j(t+\tau)}}, \qquad (8)$$

where $p(t)$ is the reference field. To achieve robustness, it is averaged over blocks of data at different time t or over signals of different frequencies, or both.

A third method to measure the signal temporal coherence uses broadband signals separated by a delay time τ,

$$\rho(t,\tau) = \left\langle \frac{[p^*(t)\otimes p(t+\tau)]_{\max}}{\sqrt{[p^*(t)\otimes p(t)]_{\max}[p^*(t+\tau)\otimes p(t+\tau)]_{\max}}} \right\rangle, \qquad (9)$$

where $[p^*(t)\otimes p(t+\tau)]_{\max}$ means the maximum value of the cross-correlation of the two time series or the convolution of the time-reversed signal (denoted by *) with the other signal. A broadband signal is chosen as the reference signal and correlated with the signals arriving at a later time. The ensemble average will be over signals transmitted at different geotime t, i.e., different reference signals. Broadband temporal correlation can be shown to be equivalent to narrow band correlations summed over frequencies. Given enough data samples, these two measures should yield the same answer, at least in theory. This (third) method will be used extensively for data analysis in this paper.

As shown by Eq. (9), the signal temporal coherence is implicitly a function of t, the signal transmission time (or geotime). The measurement (averaged over data) near time t will be the "instantaneous" signal coherence-time. The measurement of the signal coherence-time could have a large variance depending on the signal frequency and the time scale of the oceanographic processes affecting the signal. The interpretation of the coherence-time data needs to be put in the context of the corresponding propagation environment; whatever models are used needs to reproduce the measured channel impulse responses first.

III. ASCOT01 DATA ANALYSIS: THE MEASUREMENT ISSUES

In this section we analyze the data from the ASCOT01 experiments. It also addresses some of the measurement issues. For example, what are the suitable waveforms for temporal coherence measurement? What is the signal transmission rate to properly sample the temporal coherence? What is the effect of noise on the temporal coherence measurement? What is the minimum size of the samples for temporal coherence measurements? What is the effect of array motion on the measurements? Finally, how is the measurement influenced by the number of multipaths and the signal bandwidth?

The ASCOT01 experiment took place in June of 2001 off the coast of New England along the contour of 100 m depth.[21,22] It was part of joint SACLANTCEN and US experiments. The sound speed is nearly a constant with a velocity of 1465 m/s for depths below 30 m. The sound speed
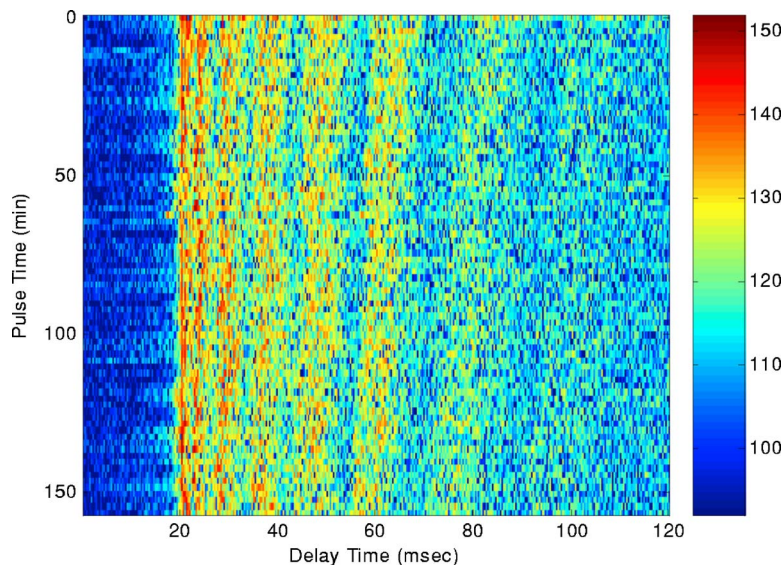
FIG. 1. (Color online) Measured channel impulse responses as a function of the delay time and geotime (time relative to the first pulse) based on LFM signals transmitted in the ASCOT01 experiment. Color scale is dB. Signal center frequency is 3550 Hz and source-receiver range is 10 km.

increases gradually from 30 m upward to 1500 m/s at 8 m depth. The source was mounted 4 m from the ocean bottom. A vertical line array was deployed at three distances of approximately 2, 5, and 10 km from the source. The data at 5 and 10 km will be used to study the range dependence of the temporal coherence of the signal. At each range, broadband phase-modulated signals were transmitted at three carrier frequencies, 450, 1200, and 3550 Hz. The data at different frequencies will be analyzed to study the frequency dependence of the temporal coherence. The signals at the lower two carrier frequencies had a bandwidth of 400 Hz. At 3550 Hz, signals were transmitted using several different bandwidths extending from 500 to 2000 Hz.

In the following subsections, the analysis focuses on data collected at the range of 10 km using three different carrier frequencies, followed by the data at 5 km. The receiver depth is between 65 and 82 m depth. No significant depth dependence is found in the temporal coherence result.

The signals analyzed were acoustic communication signals, transmitted in packets. Each packet lasted about 10 or 20 s depending on the signal frequency. A packet started with a LFM signal of 0.1 s, followed by a 0.3 s gap, followed by a Barker code signal of 13 symbols, followed by a 0.3 s gap, and then followed by the data signal. For the temporal coherence study, the data were BPSK signals coded with repetitive m-sequences. The m-sequences are "orthogonal" signals such that the signal autocorrelation is sharply peaked with a peak-to-sidelobe ratio of $10 \log N$, where N is the number of symbols used in the sequence. Unless otherwise mentioned, the signal packets were repeated every 2 min over a period of several hours. The LFM signal has the same bandwidth as the m-sequence signal.

LFM signals have been commonly used in the past to measure the channel impulse response functions and estimate the signal temporal coherence. To allow the reverberation to die down, the pulses were usually transmitted on the order of minutes to tens of minutes. It should become clear below that for channels with a short coherence-time, the LFM signal is a poor choice as the channel is too infrequently sampled. This is particularly true at high frequencies. To sample the

channel at a high sampling rate, m-sequence signals are used, transmitted in sequence continuously for an extended period of time. The m-sequence signals sample the channel impulse responses at a period of N/bandwidth s.

A. 3550 Hz at 10 km range

The channel impulse responses were obtained by matched filtering the data with the transmitted signal. The channel impulse responses from the LFM signals are plotted in Fig. 1 displaying the multipath arrivals as a function of delay time and geotime (transmission time); the LFM signals were transmitted every 2 min. The signal bandwidth is 2 kHz. One clearly sees the variation of the channel impulses between packets. One notes in Fig. 1 eight groups of multipath arrivals, with each group consisting of many micropath arrivals. The micropath arrivals are rays that scatter from the ocean inhomogeneities producing microrays. As a result, the arrival times of the micropath arrivals are spread out and their intensities fluctuate with time. The experimental site (in the New England coastal water) is known to have internal waves and turbulence.[22] The effect of these oceanographic processes on the acoustic signal propagation is evident in Fig. 1, particularly the late arrivals.

(One notes a gradual decrease in the delay time in Fig. 1 between the early and late arrivals as geotime increases. The shift in the arrival time of the late arrivals can be explained in terms of the internal tides, which shift the depth of the channel thermocline.[23] See Sec. VI D for further details).

Take one impulse response function as a reference and cross correlate with the impulse responses transmitted at a later time. The maximum values of the correlations are entered in Eq. (9) and one obtains one temporal coherence curve, as shown in Fig. 2(a). Take different impulse response functions as the references; one obtains a distribution of temporal coherence curves at different geotimes, as shown in Fig. 2(a). The average of all the curves is shown in Fig. 2(b), displayed by the data points marked as crosses. One observes that the average coherence decreases to a value of ~ 0.25 over a period of 5–20 min.
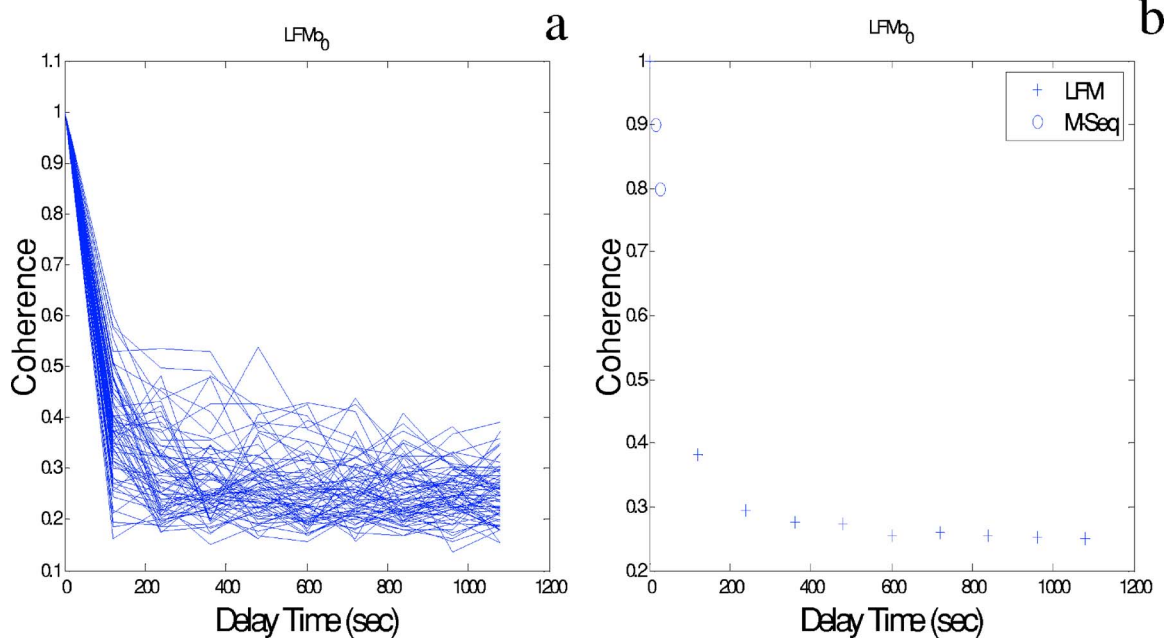
FIG. 2. (Color online) Temporal coherence curves measured at a different starting time (a), and averaged over 2.5 h of data using LFM signals (b). This figure shows that the normal method using LFM pulses is inadequate for measuring the temporal coherence at a short delay time. Superimposed in Fig. 4(b) is the measurement from m-sequence data from Fig. 4. Center frequency=3550 Hz, and range=10 km.

Figure 2(a) shows a wide spread in time for a fixed coherence value. For example, for a coherence value at $1/e$, the coherence-time can vary from 120 s to greater than 1000 s. Thus, the conventional definition of coherence-time, defined by the time delay at which the coherence value drops to $1/e$, could not be applied directly to this data. In general, the variance of the coherence-time at a fixed coherence value increases significantly as the coherence value decreases. In this paper, the measurement of the signal coherence-time is based on the time when the coherence drops to 0.8, i.e., $\tau_{0.8}$. The reason is to keep the STD of the (coherence-time) measurements smaller than the average (coherence-time). The coherence-time $\tau_{1/e}$ can be extrapolated from $\tau_{0.8}$ given a model of temporal coherence.

Figure 2(b) shows that the coherence value drops to be-low 0.5 at the first sample, the 120 s delay time. To measure $\tau_{0.8}$, it is necessary to increase the sampling rate using a shorter sampling interval. To this end, the m-sequence data were analyzed next. Each m sequence contained $N=1023$ symbols and lasted 0.51 s. The m sequences were repeated for a total duration of 17 s.

The m-sequence data were processed the same way as the LFM signals. Figure 3 shows the channel impulse response for m sequences for one packet of data. Because the m-sequence signals are nearly orthogonal, with a sidelobe level that is, in principle, $10 \log N$ dB down from the main level, the (circular) matched-filter correlation of the m-sequence data yields a good measurement of the channel impulse response function. Note that the amplitudes and arrival time are pretty stable for multipath arrival with a delay
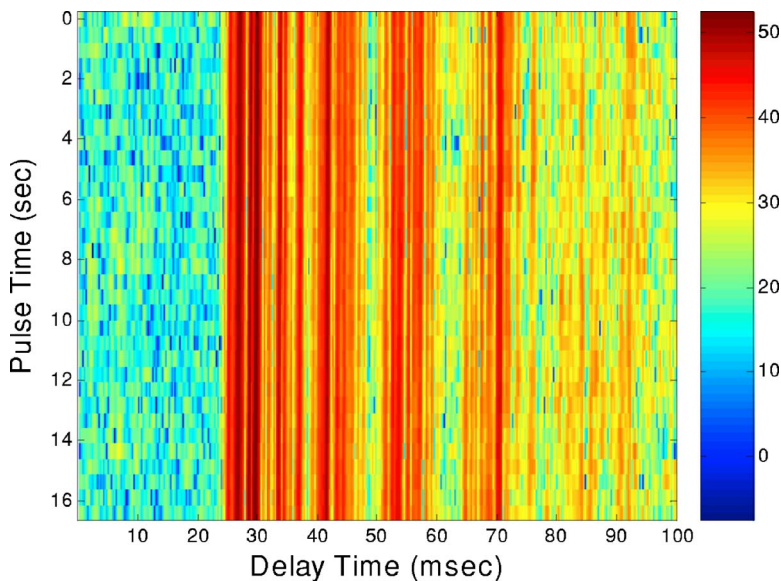
FIG. 3. (Color online) A typical display of the channel impulse responses measured from consecutive m sequences within a packet as a function of the delay time and geotime. Center frequency is 3550 Hz and range is 10 km. Color scale is dB. The m sequences are commonly used to characterize the channel variation within a packet of acoustic communication signal.
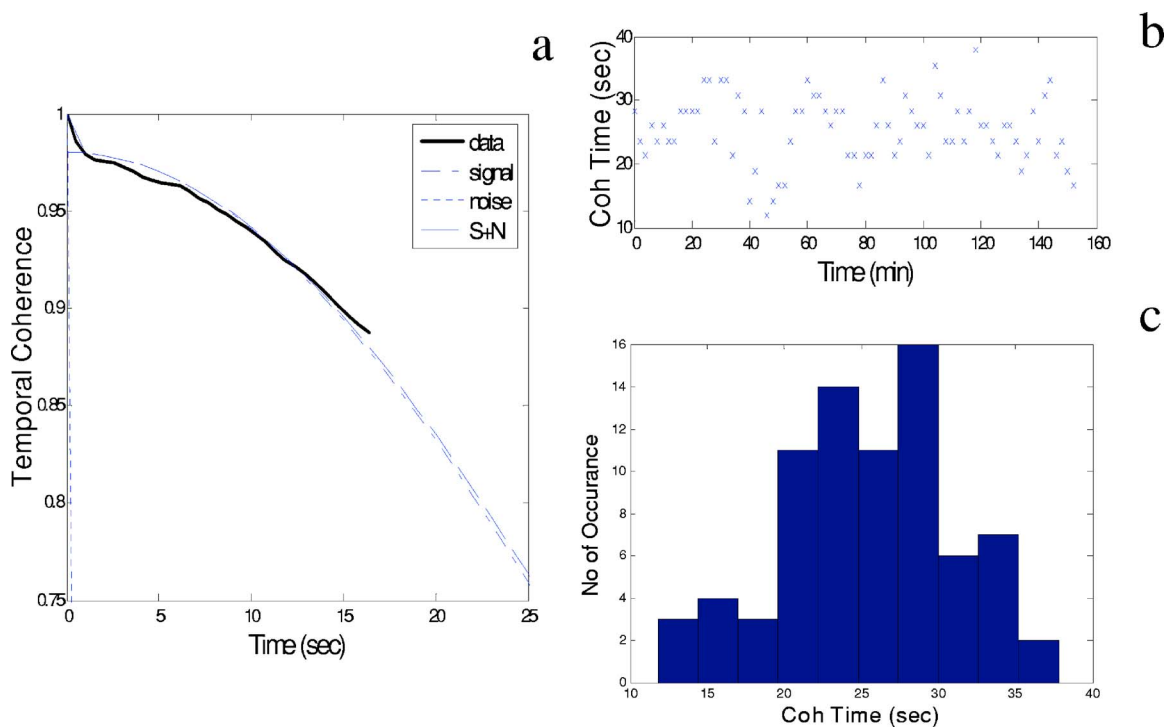
FIG. 4. (Color online) Average temporal coherence of the m-sequence signals, compared with a model of signal and noise (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c). Center frequency=3550 Hz, range=10 km.

time of less than 50 ms. Significant amplitude fluctuations are noted for arrivals with a delay time greater than 50 ms. This implies that the m-sequence signals are highly coherent within a packet.

Figure 4(a) shows the temporal coherence of the m sequences averaged over 77 packets of data (the solid curve), covering a time span of 2.5 h. Note a sharp decrease of the measured temporal coherence near the origin with a slope very different from the rest of the curve. To explain this, one needs to include the contribution of noise in Eq. (1). Denoting the signal as s and noise as n, thus $p=s+n$. The measured temporal coherence $\rho(\tau)$ is found to be

$$\rho(\tau) = s(\tau)\frac{S^2}{S^2 + N^2} + n(\tau)\frac{N^2}{S^2 + N^2}, \qquad (10)$$

where $s(\tau)$ is the temporal coherence of the signal and $n(\tau)$ is the temporal coherence for the noise. In deriving Eq. (10), it is assumed that the signal and noise are uncorrelated. The coefficients multiplying the signal and noise coherence terms in Eq. (10) are the signal to signal-plus-noise ratio and noise to signal-plus-noise ratio.

One notes that noise is normally assumed uncorrelated in the temporal space. It has in reality a very short coherence-time. The sharp drop of the measured coherence in Fig. 4(a) at a short delay time can be attributed to the noise, as illustrated later. The signal is expected to have a coherence-time much longer than the noise. Consequently, the signal coherence curve should have a slower (absolute) slope than the noise coherence curve. The data with a delay time greater than 1 s are that associated with the signal. One can fit the data in Fig. 4(a) using Eq. (10), assuming that both the signal and noise coherence are modeled by Eq. (6). The

modeled signal and noise coherence curves are shown in Fig. 4(a) as the dash-dotted and dotted lines, respectively, assuming $\tau_{lle}=50$ and 0.1 s for the signal and noise, respectively. (The signal coherence curve is displaced by 0.02 from unity.) The sum of the signal plus noise coherence using Eq. (10) is shown as the dashed line and is found in good agreement with the measured coherence data (solid curve). In fitting the data, one assumes a value of 0.02 for the noise to signal-plus-noise ratio, which is estimated by the contribution of the second term in Eq. (10) at zero delay time. This yields a signal-to-noise ratio of 17 dB, which agrees with the estimation based on the raw data.

The data analysis procedure is to determine the signal coherence from the measured coherence data by removing the noise contribution, which concentrates at a delay time less than 1 s. From the signal temporal coherence, one then estimates the signal coherence-time. From the data shown in Fig. 4(a), one finds the coherence-time of the average signal coherence at a coherence value of 0.8 is about 23 s. One can also obtain the "instantaneous" coherence-time of the signal by applying the above procedure to the temporal coherence curve measured from each packet of the m-sequence data. The measured "instantaneous" coherence-time of the signal is plotted in Fig. 4(b) as a function of geotime of the transmitted packet. Figure 4(c) shows the histogram distribution of the "instantaneous" coherence-time of the signal. From the histogram one obtains a mean coherence-time $\tau_{0.8}$ of around 25 s, approximately the same as the coherence-time of the average signal coherence reported above. One also finds a STD of ~5 s for the mean coherence-time measurement.

The above data were also used to study the bandwidth
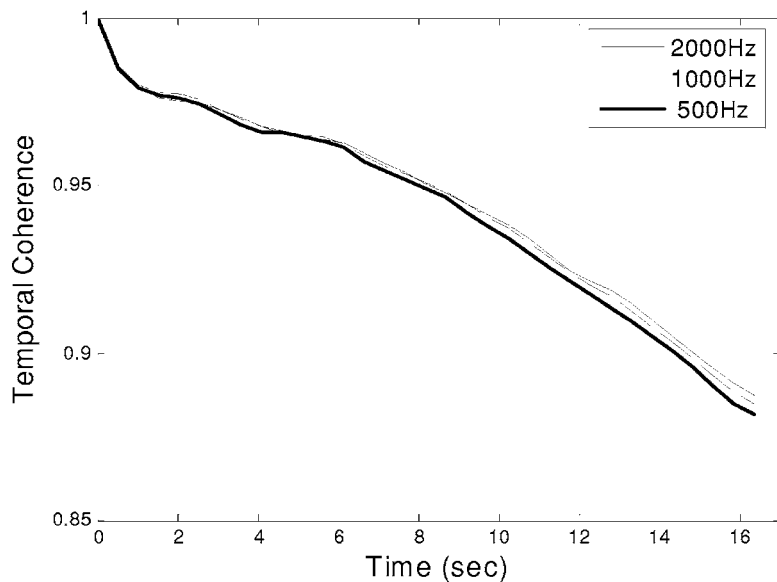
FIG. 5. Bandwidth dependence of the average temporal coherence of the signal. Center frequency=3550 Hz, range=10 km.

dependence of the signal temporal coherence. In general, the signal coherence will depend on the signal bandwidth; a wideband signal will have a different temporal coherence than a narrow band signal. In addition to sensing different frequency responses of the propagation channel, a wide band signal can resolve individual (group of) path arrivals and measure the temporal coherence of these paths, which a narrowband signal cannot. To conduct the study, signals of different bandwidths are obtained by applying a bandpass filter to the raw data (mentioned above) using a different filter bandwidth. Temporal coherence is next deduced from the filtered data. Figure 5 shows the measurements of the temporal coherence of the bandpassed signals. One notes very little dependence of the temporal coherence on the signal bandwidth. This result is perhaps expected when the bandwidth of the signal is small compared with the signal carrier frequency, as is the case for the 3550 Hz data. The lack of bandwidth dependence as demonstrated by the 3550 Hz signal will be assumed to be true for signals at other frequencies. This allows us to compare and interpret data of different

bandwidths as presented later without considering the bandwidth as a separate parameter in the equation.

B. 1200 Hz at 10 km range

The channel impulse responses measured from the LFM signals are plotted in Fig. 6. The signals were transmitted every 2 min covering a period of 8 h. The LFM signals had a bandwidth of 400 Hz. Consequently, the temporal resolution in Fig. 6 is not as good (sharp) as that seen in Fig. 1. Nonetheless, one still sees eight groups of multipath arrivals. Figure 6 shows clearly that the multipath arrival time (particularly the last four groups of multipaths) also varies with geotime. The gradual shift of the late arrivals with geotime is similar to that shown in Fig. 1 and can be explained by the shift of the thermocline axis due to (semidiurnal) internal tides.

The impulse responses at a different geotime were used as the references and cross-correlated with the impulses responses that follow. For each reference, a temporal coher-
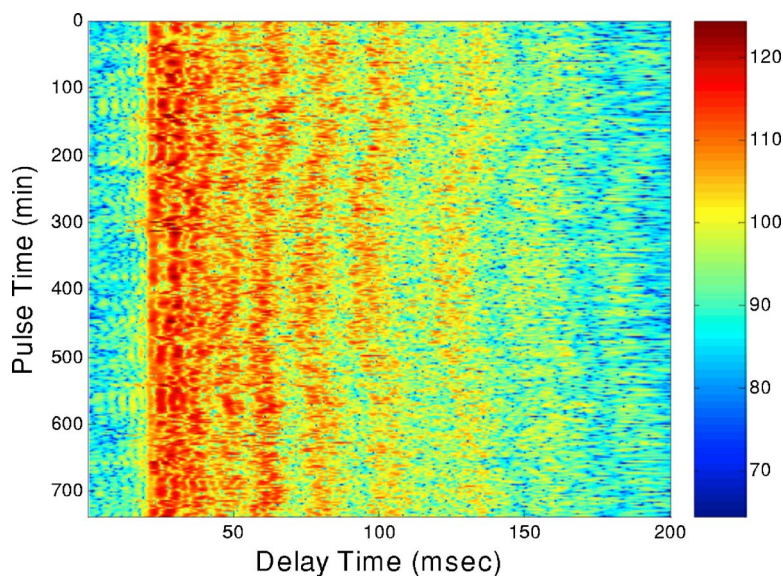
FIG. 6. (Color online) Channel impulse responses at 1200 Hz as a function of the delay time and geotime, based on LFM data received at a range of 10 km during the ASCOT01 experiment. Color scale is dB. The multipath arrival pattern is similar to that shown in Fig. 1.
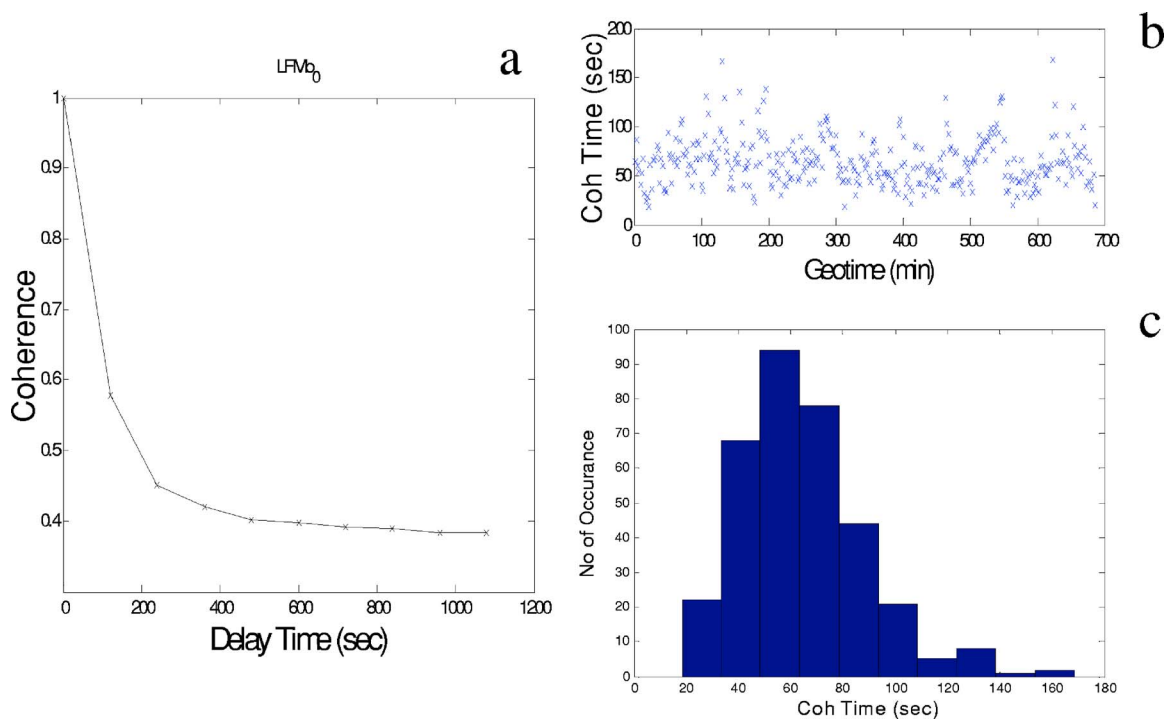
FIG. 7. (Color online) Temporal coherence measured from data shown in Fig. 6. Average temporal coherence as a function of signal separation time (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c).

ence curve was obtained. Figure 7(a) shows the average of all the temporal coherence curves. From this, one can estimate the coherence-time of the average signal coherence to be $\tau_{0.8} \simeq 57$ s. One can also estimate the coherence-time $\tau_{0.8}$ for each of the "instantaneous" temporal coherence curves obtained. The results are plotted in Fig. 7(b) to show the temporal variation (time dependence) of the "instantaneous" signal coherence-time. It indicates a cycling period of 30–50 min, which is shorter than the periods of the internal tides. The cause of the temporal variation of the signal coherence-time has not been investigated.

Figure 7(c) shows the histogram of the measured coherence-time. One finds a mean of approximately 64 s and a STD of about 24 s.

C. 450 Hz at 10 km range

The 450 Hz signals had a bandwidth of 400 Hz covering a frequency range of 250 to 650 Hz. Figure 8 shows the measured channel impulse responses deduced from the LFM signals. The signals, transmitted every 2 min, covered a period of 3.5 h. The impulse responses at 450 Hz are clearly different from the impulse responses shown earlier at 1200 or 3550 Hz. Although seven to eight groups of multipath arrivals can also be identified in Fig. 8, one finds that the second and fifth multipaths are the dominant arrivals versus the first to third arrivals in Figs. 1 and 6.

The impulse responses were processed to produce the temporal coherence curves at a different geotime (as de-
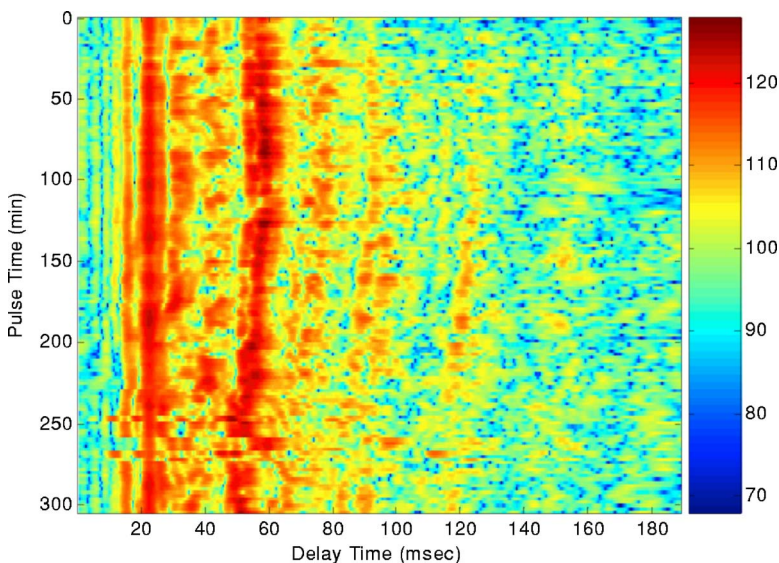
FIG. 8. (Color online) Channel impulse responses at 450 Hz as a function of the delay time and geotime based on LFM signals transmitted during the ASCOT01 experiment. Color scale is dB. Source-receiver range is 10 km. Second and fifth arrivals have higher amplitudes than the others. Impulse responses are aligned by the peak amplitude within the first 30 ms of data; data at geotime from 240 to 270 min are misaligned by this method.
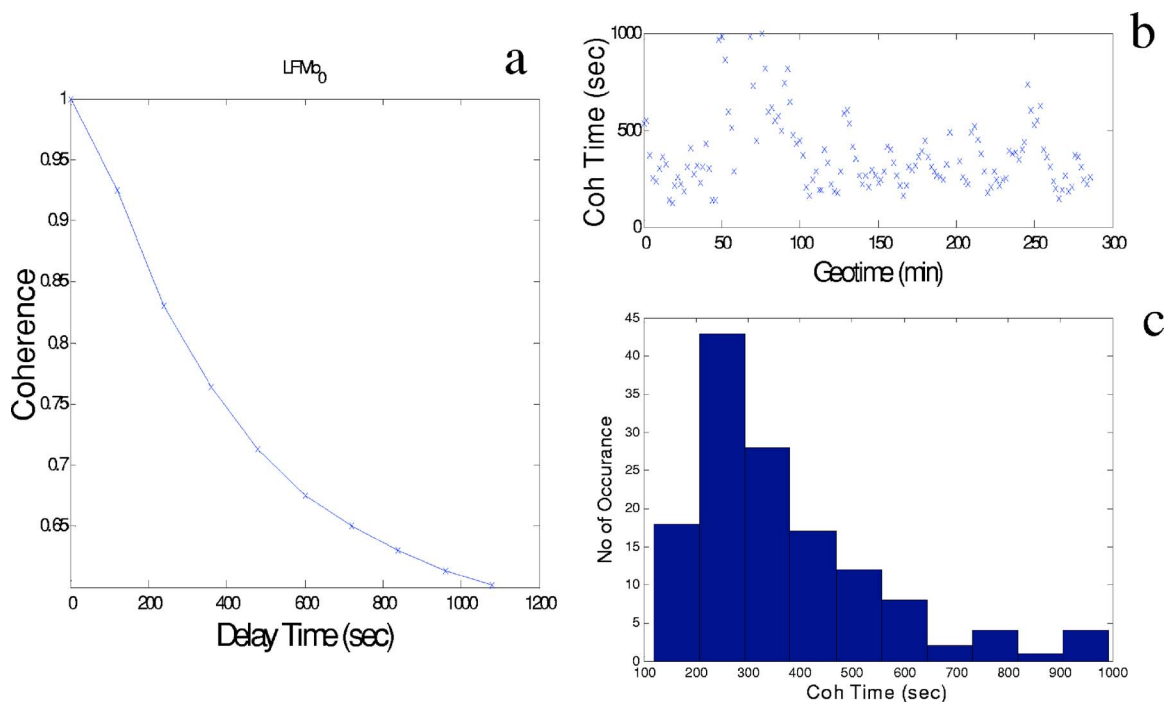
FIG. 9. (Color online) Temporal coherence based on data of Fig. 8. Average temporal coherence as a function of signal separation time (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c).

scribed above). Figure 9(a) shows the average of all the temporal coherence curves. One finds the coherence-time $\tau_{0.8}$ of the average signal coherence to be about 299 s. The coherence-time estimated from each of the temporal coherence curves (described above) is plotted in Fig. 9(b). Figure 9(b) shows the variation (time dependence) of the "instantaneous" signal coherence-time. Figure 9(c) shows the histogram of the measured coherence-time. One finds a mean $\tau_{0.8}$ of approximately 382 s and a STD of about 202 s. The high value of the STD arises from the high "instantaneous" coherence-time during the period of 50–90 min in Fig. 9(c), which is distinctly different from the rest of the time. There is no apparent correlation between the time series of the "instantaneous" coherence-time [Fig. 9(b)] and the multipath arrival time (Fig. 8); the later is related to the tidal cycle.

Note that the STD measured over a short period of time (10–40 min) is much smaller ($\sim$100 s). For a smaller time period (minutes), the STD will be smaller. Thus the STD in the coherence-time could be used as an indicator of the amount of the changes in the acoustic environment (e.g., the sound speed STD) during the measurement periods. See Sec. VI for further study.

D. Data at 5 km range

The data at the 5 km range were processed in the same manner as the data at the 10 km range. Only LFM signals were used. The results for 3550, 1200, and 450 Hz are shown in Figs. 10–15, respectively.

The measured channel impulse responses at 3550 Hz are
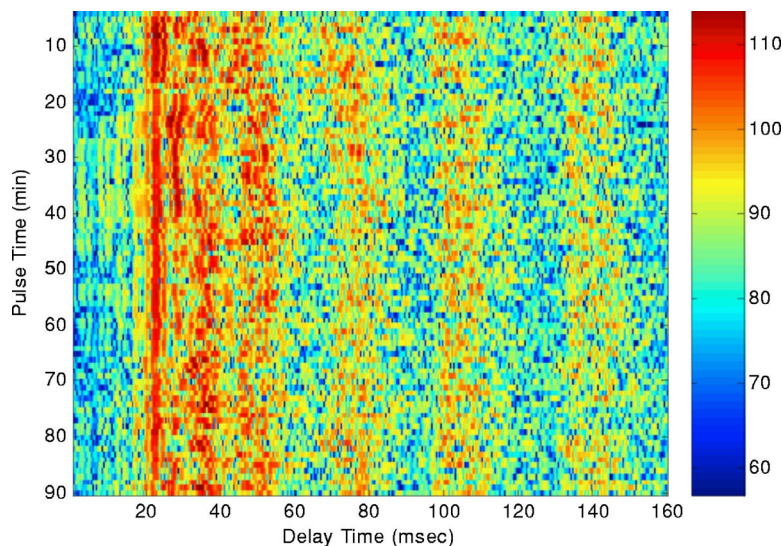
FIG. 10. (Color online) Channel impulse responses at a range of 5 km as a function of the delay time and geotime based on LFM signals centered at 3550 Hz from the ASCOT01 experiment. Color scale is dB. Similar arrival pattern as in Fig. 1 but with a nearly doubled multipath spread.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Tsih C. Yang: Temporal coherence of acoustic signals   2603
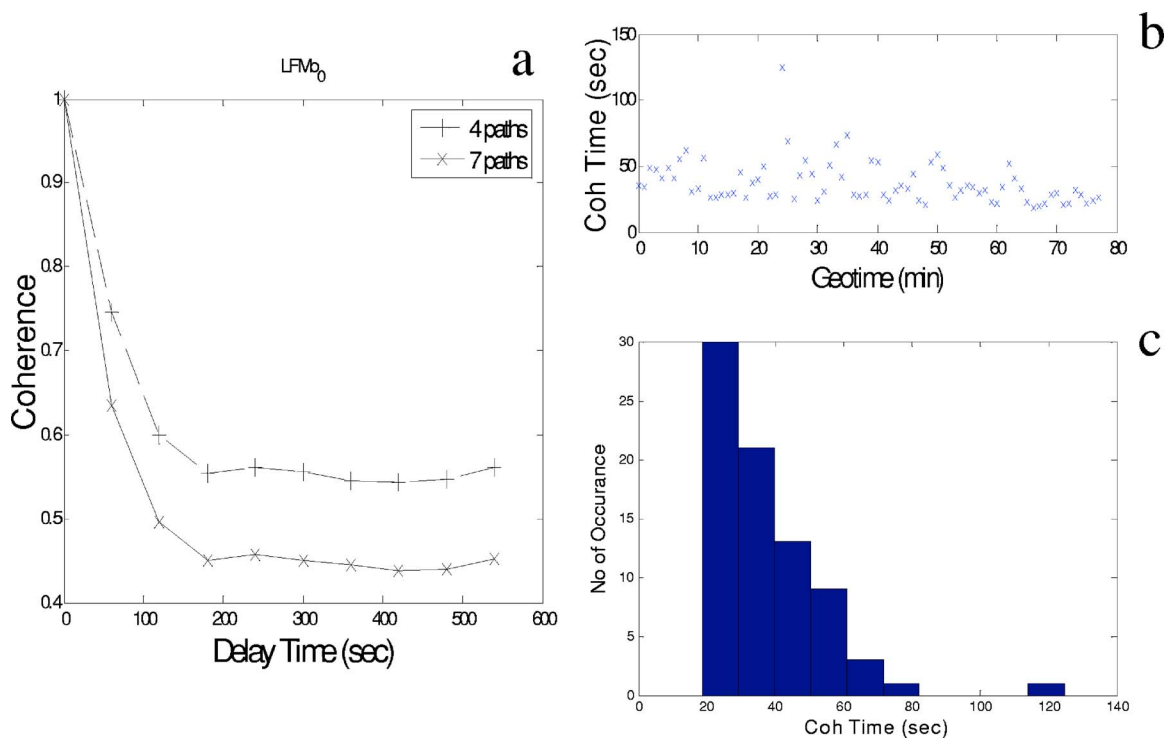
FIG. 11. (Color online) Temporal coherence measured from date of Fig. 10. Average temporal coherence using the impulse responses containing all seven paths versus only the initial four paths (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c). Center frequency=3550 Hz, range=5 km.

shown in Fig. 10 as a function of the delay time and geotime. The difference between this data (Fig. 10) and the data at 10 km (Fig. 1) is that the 3550 Hz signals at 5 km had only a bandwidth of 500 Hz. The signals, transmitted every 1 min, covered a period of 1.5 h. The channel impulse responses, deduced from the LFM signals at 1200 Hz, are plotted in Fig. 12. The signals, transmitted every 2.5 min, covered a period of 4 h. The channel impulse responses deduced form the LFM signals at 450 Hz are plotted in Fig. 14. The signals, transmitted every 2.5 min, covered a little more than 3 h. Both the 1200 and 450 Hz signals have a 400 Hz bandwidth, the same as the 10 km case.

One observes seven to eight groups of multipaths in

Figs. 10, 12, and 14. Certain groups of arrivals are clearly split into two subgroups of arrivals in Figs. 12 and 14. Comparing Figs. 10, 12, and 14 with Figs. 1, 6, and 8, one finds that the spreads of the multipath delay times (from the first arrival) at 5 km are almost double that at 10 km. The peak intensity distribution of the 450 Hz impulse responses changes significantly from 10 to 5 km.

The impulse responses were processed (as described above) to produce the temporal coherence curves at a different geotime—using the impulse response at a different time as the reference. coherence-time $\tau_{0.8}$ was estimated from each of the temporal coherence curves.

Figure 11(a) shows the average temporal coherence of
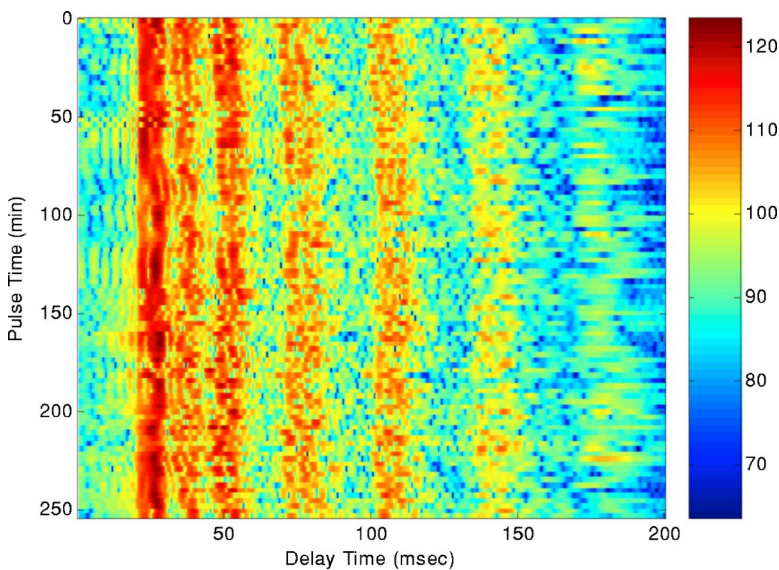
FIG. 12. (Color online) Channel impulse responses as a function of the delay time and geotime at a range of 5 km and center frequency of 1200 Hz based on LFM signals from the ASCOT01 experiment. Color scale is dB. Similar multipath arrival pattern as in Fig. 6, but with a nearly doubled delay spread.
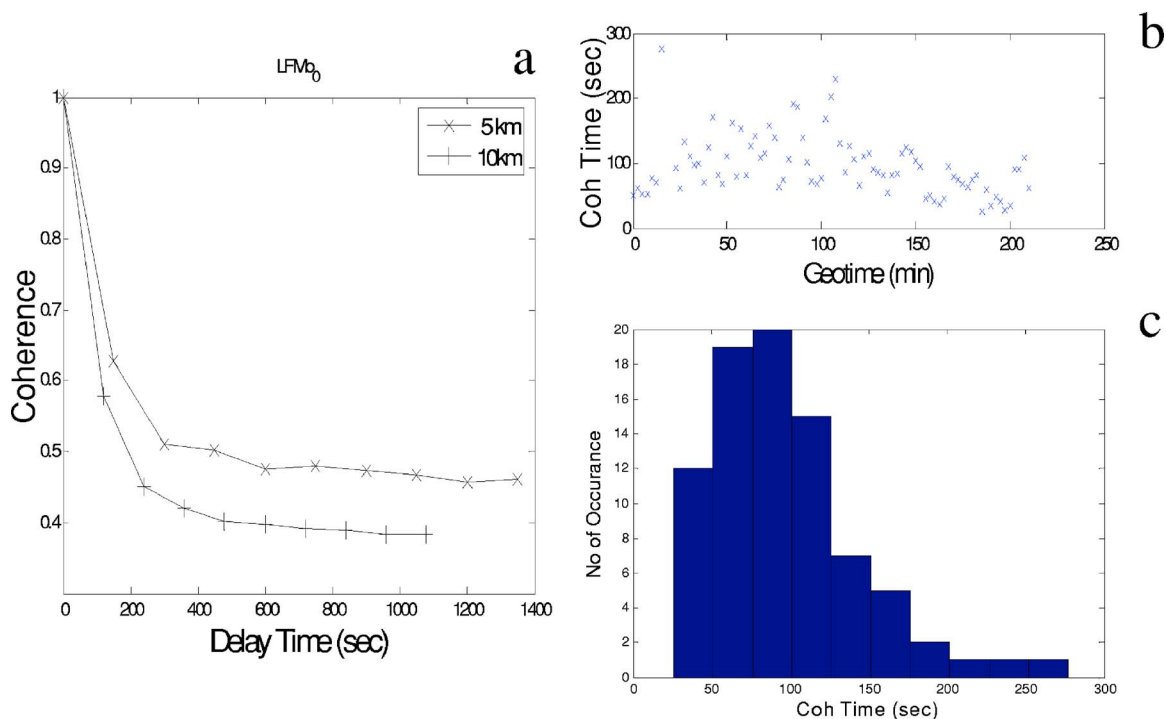
FIG. 13. (Color online) Temporal coherence measured from the data of Fig. 12. The average temporal coherence as a function of the signal separation time (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c). The average temporal coherence of the signal at 10 km [Fig. 7(a)] is also displayed in Fig. 13(a) for comparison.

the signal at 3550 Hz (the solid curve). One finds the coherence-time $\tau_{0.8}$ of the average signal coherence to be about 33 s. Figure 11(b) shows the temporal variation (time dependence) of the "instantaneous" signal coherence-time at 3550 Hz. Figure 11(c) shows the histogram of the measured coherence-time. One finds a mean of approximately 37 s and a STD of about 16 s for the 3550 Hz signal at 5 km.

The signal coherence-time of the 1200 Hz signal at 5 km is shown in Fig. 13. Figure 13(a) shows the average of all the temporal coherence curves (the solid curve) at 5 km. One finds the coherence-time $\tau_{0.8}$ of the average signal coherence to be about 81 s. Superimposed is the previously obtained average temporal coherence curve at 10 km (the

dashed curve). Figure 13(b) show the temporal variation (time dependence) of the "instantaneous" signal coherence-time at 1200 Hz. Figure 13(c) shows the histogram of the measured coherence-time. One finds a mean of approximately 96 s and a STD of about 46 s.

The signal coherence-time of the 450 Hz signal at 5 km is shown in Fig. 15. Figure 15(a) shows the average of all the temporal coherence curves (the solid curve). One finds the coherence-time $\tau_{0.8}$ of the average signal coherence to be about 443 s. Superimposed is the previously obtained average temporal coherence curve at 10 km (the dashed curve). Figure 15(b) shows the temporal variation (time dependence) of the "instantaneous" signal coherence-time at 450 Hz. Fig-
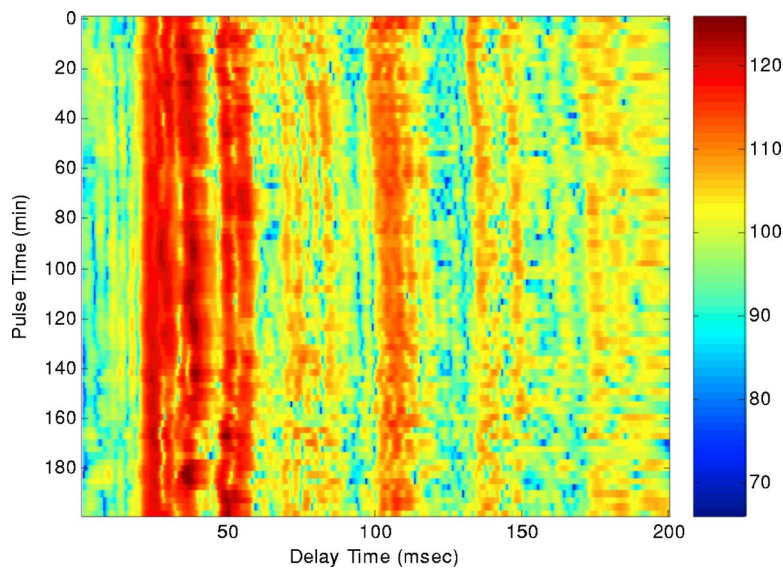
FIG. 14. (Color online) Channel impulse responses as a function of the delay time and geotime at a range of 5 km and center frequency of 450 Hz based on LFM signals from the ASCOT01 experiment. Color scale is dB. Multipath delay spread is almost twice that of Fig. 8, but the amplitude distribution is different.
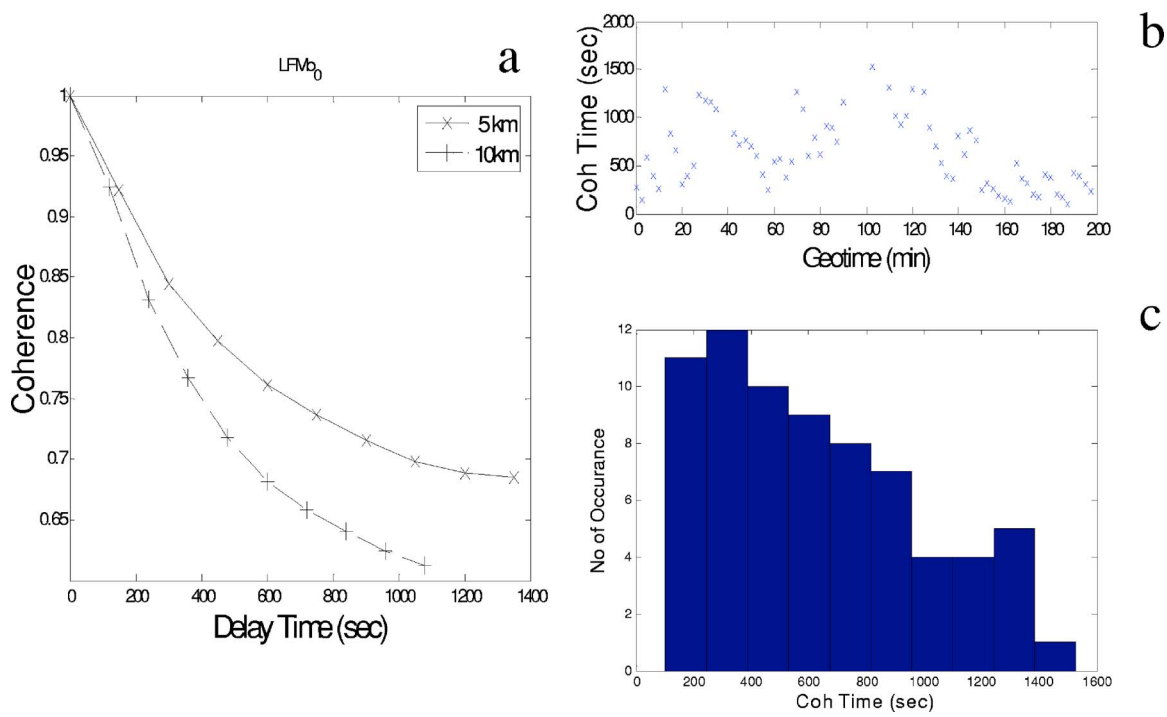
FIG. 15. (Color online) Temporal coherence measured from data of Fig. 14. Average temporal coherence of the LFM signals as a function of signal separation time (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c). The average temporal coherence of the signal at 10 km [Fig. 9(a)] is also displayed in Fig. 15(a) for comparison.

ure 15(c) shows the histogram of the measured coherence-time. One finds a mean of approximately 626 s and a STD of about 368 s.

E. Comments on the coherence measurements

As is well known, the temporal coherence of the signal may vary from path to path. In general, the early arrivals are more coherent than the latter (scattering) arrivals. Hence, measurement of the temporal coherence could depend on the number of multipaths included in the signal. To illustrate this, it is plotted in Fig. 11(a) the average temporal coherence of the signal using only (the impulse responses of) the first four multipaths (the dashed curve) to compare with that using all seven multipaths (solid curve). One finds that the temporal coherence value increases by ~20% at elapsed time >100 s.

The measurements of the signal temporal coherence in this paper include all the "dominant" multipath arrivals. To do so, one must use a sufficiently long time window for measuring the impulse response delay spread, yet not too long to increase the noise contribution. The procedure used here is to increase the window until the (average) coherence value converges to within a certain tolerance. The final time-window-used is the minimum time window-that achieves the convergence value.

One observes that the mean of the coherence-times is generally somewhat higher than the coherence-time estimated from the average of the temporal coherence curves. Note that the former is the average of the ("instantaneous") coherence-time at a fixed coherence value, whereas the latter is deduced from the average of the coherence at fixed elapsed times. Note that the measurements are only available at dis-

crete elapsed times. Hence measurement data are interpolated or extrapolated to determine the coherence-time $\tau_{0.8}$. The details of interpolation and extrapolation, which involves curve fitting, could contribute to the difference between the two different ways of measuring the coherence-time, particularly when the "instantaneous" temporal coherence exhibit high degrees of fluctuations. It is observed in Fig. 2(a) that the variance of the data at a fixed coherence value is much higher than the variance of the data at a fixed elapsed time. Consequently, it is assumed that the coherence-time estimated from the average of the temporal coherence curves is more trustworthy than the mean of the coherence-times. The coherence-time estimated from the average of the temporal coherence curves is entered as the coherence-time $\tau_{0.8}$ in Table I. The STD measured above, which represents the spread (or the temporal variation) of the "instantaneous" coherence-time, is also entered in Table I. This STD represents the environmental influence on the signal coherence-time. It should not be confused with the STD of the measurements (or the measurement error due, for example, to the different sizes of data samples). Note that the measurements of the temporal coherence for a broadband signal use an

TABLE I. coherence-time $\tau_{0.8}$ and the spread of the coherence-time (in parentheses) measured from ASCOT01 data; the later reflects the amount of the temporal variation of the coherence-time due to changes in the propagation environment.

Frequency	$\tau_{0.8}$ at 10 km	$\tau_{0.8}$ at 5 km
3550 Hz	23 (±5) s	33 (±16)
1200 Hz	57 (±24) s	81 (±46)
450 Hz	299 (±202) s	443 (±368)
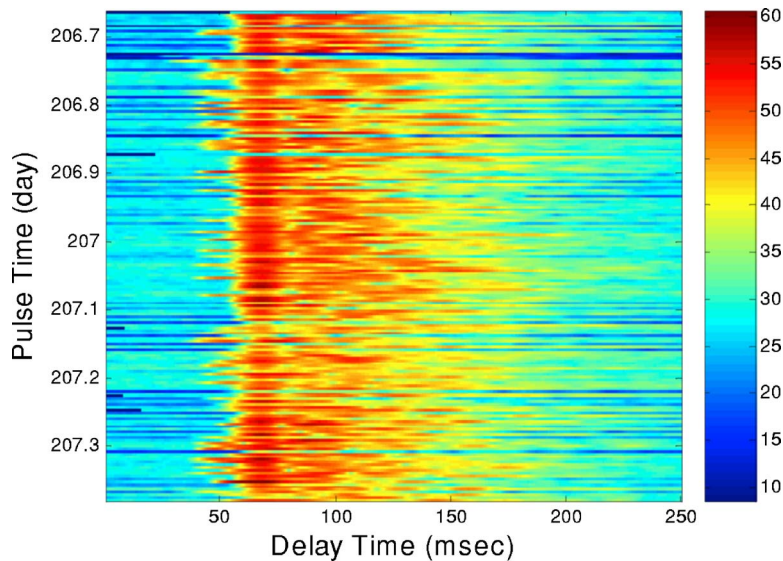
FIG. 16. (Color online) Results from the SWARM95 experiment. Measured channel impulse responses as a function of the delay time and geotime, showing groups of arrivals. Color scale is dB. Data based on m-sequence signals centered at 400 Hz with a range at 42 km.

average over frequencies. In contrast, the temporal coherence for a narrow band signal is averaged over data at different time. Both use an adequate number of samples.

IV. SWARM 95 DATA ANALYSIS

In this and the next section, the above analysis is applied to data from another site, off the coast of New Jersey. Although the measurement procedures are the same and repetitive as before, the data features (the delay structure and time variations of the channel impulse responses) are very different. How is signal coherence-time related to the channel dynamics is of great interest and will be briefly discussed in Sec. VI.

The SWARM 95 experiment was conducted in July–August of 1995 in the Mid-Atlantic bight continental shelf region east of New Jersey. It was sponsored by the Office of Naval Research with participation from many institutions. Details of the experiment were described in a previous publication.[24] The data to be analyzed here are data from two acoustic sources located upslope in 55 m of water to the NRL vertical line array down slope in 88 m of water. The source receiver range was ~42 km. A 400 Hz source was deployed at midwater depth and a 224 Hz source was deployed near the bottom. The 400 Hz signals had a bandwidth of 100 Hz and the 224 Hz signals had a bandwidth of 16 Hz. The NRL vertical line array (VLA) contained 32 elements spanning the water column from 21 to 85 m.

BPSK data using m sequences were analyzed following the analysis described above. The BPSK signal at 400 Hz used an m sequence of 511 symbols repeated 23 times. Each packet lasted 117.53 s and was repeated continuously with a gap of 242.47 s in between packets. The BPSK signal at 224 Hz used an m sequence of 63 symbols repeated 30 times. Each packet lasted 118.13 s and was repeated continuously with a gap of 181.9 s in between packets.

A total of 179 packets was analyzed for the 400 Hz signal, yielding 4117 measurements of channel impulse responses covering a period of 17.9 h. The channel impulse responses were plotted in Fig. 16 for a receiver at 51 m depth. Temporal coherence was estimated for each packet

using the first m sequence as the reference. The average of all the coherence curves is plotted in Fig. 17(a). It yields a coherence-time $\tau_{0.8} \simeq 40$ s. The "instantaneous" coherence-time estimated from each packet is shown in Fig. 17(b). Figure 17(c) shows the histogram distribution of the "instantaneous" coherence-time. One finds a mean coherence-time of 55 s and STD of 40 s. The high STD is a manifestation of the (significant) environmental change in the SWARM95 area due to the presence of internal waves.

For the same period, the 224 Hz data yield 6450 measurements of channel impulse responses, which are plotted in Fig. 18(a). It showed one dominant group of arrivals. The mean temporal coherence averaged over all packets is plotted in Fig. 18(b), from which one finds the coherence-time $\tau_{0.8} \simeq 130$ s. In this case, one finds significantly larger fluctuations in the measurements of the "instantaneous" coherence-time than in the case of 400 Hz signal. For many packets, the coherence-time is beyond the packet length of 118 s and cannot be reliably estimated.

The SWARM95 experiment had extensive oceanographic measurements, which were carried out almost concurrently with the acoustic measurements.[24] The oceanographic data suggested the presence of solitary internal waves in addition to diffuse internal waves. The solitary internal waves are generated on the shelf edge by interaction of internal tides with the shelf break. These waves occurred on a diurnal cycle. Hence, our analysis was carried over 17 h of data to include the effect of the solitary internal waves. Internal waves are expected to decrease the signal coherence-time. The 400 Hz data indicated a higher standard deviation to the mean ratio than other experimental data (see Table I). The large time variation is presumed to be caused by the internal waves.

V. RAGS03 DATA ANALYSIS

The RAGS03 experiments contained two parts: one conducted in September 2003 for underwater acoustic communications research, and the other in the winter of 2003 for array performance study. The data of interest were from the first part of the experiment that took place in the same area as
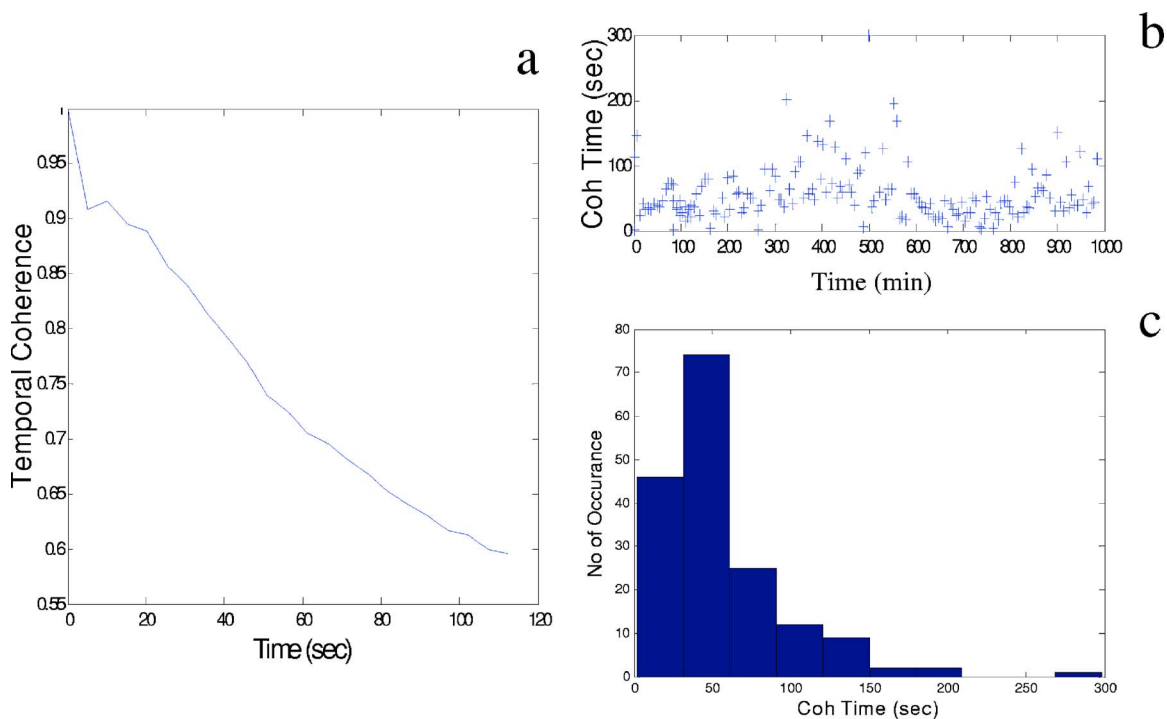
FIG. 17. (Color online) Temporal coherence measured from the *m*-sequence data of Fig. 16. Average temporal coherence as a function of signal separation time (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c).

SWARM95. The measured sound speed profiles are similar to that found in the SWARM95 experiment; both have the summer profile. There were no concurrent oceanographic measurements with the acoustic measurements as in the SWAR95 experiment, except for CTD casts. Backscattering sonar data reveal the existence of solitary internal waves similar to that found in the SWARM95 experiment.

The source was deployed at a depth of approximately 38 m projecting a broadband signal centered at 20 kHz. The receiver contained 8 hydrophones within an aperture of 2 m, at a mean depth around 41 m. Both were moored from the ocean bottom. The mean water depth is about 70 m. The

source and receivers were separated by 3 km. BPSK signals with a bandwidth of 2 kHz were used to deduce the channel impulse responses. Each packet contained ~10 s of data consisting of 19 *m* sequences each with 1023 symbols.

Figure 19(a) shows the matched filtered output of the *m*-sequence data for one packet using the transmitted *m*-sequence signal. One observes that the amplitudes of the impulse responses vary significantly between the received *m* sequences. Figure 19(b) shows the Doppler shift measured from data of individual *m* sequences determined by the peak of the broadband ambiguity function as a function of Doppler frequency and delay time. One sees a substantial varia-
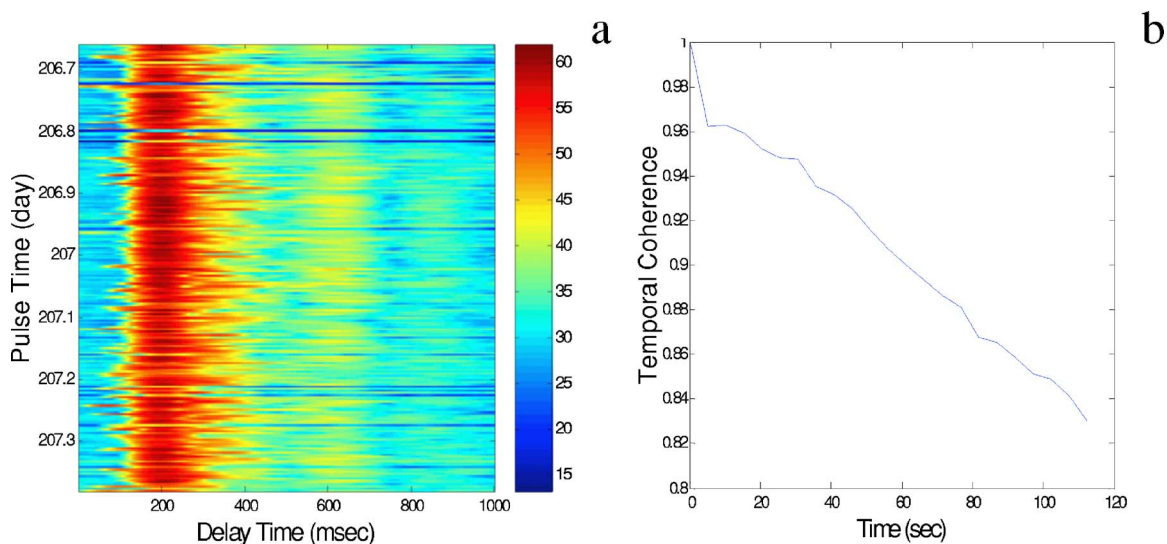
FIG. 18. (Color online) SWARM95 data 224 Hz at a range of 42 km. Channel impulse responses as a function of the delay time and geotime based on *m*-sequence signals (a). Multipath arrivals are not resolved due to the limited bandwidth of the signal (16 Hz). Color scale is dB. Average signal coherence as a function of geotime (b).
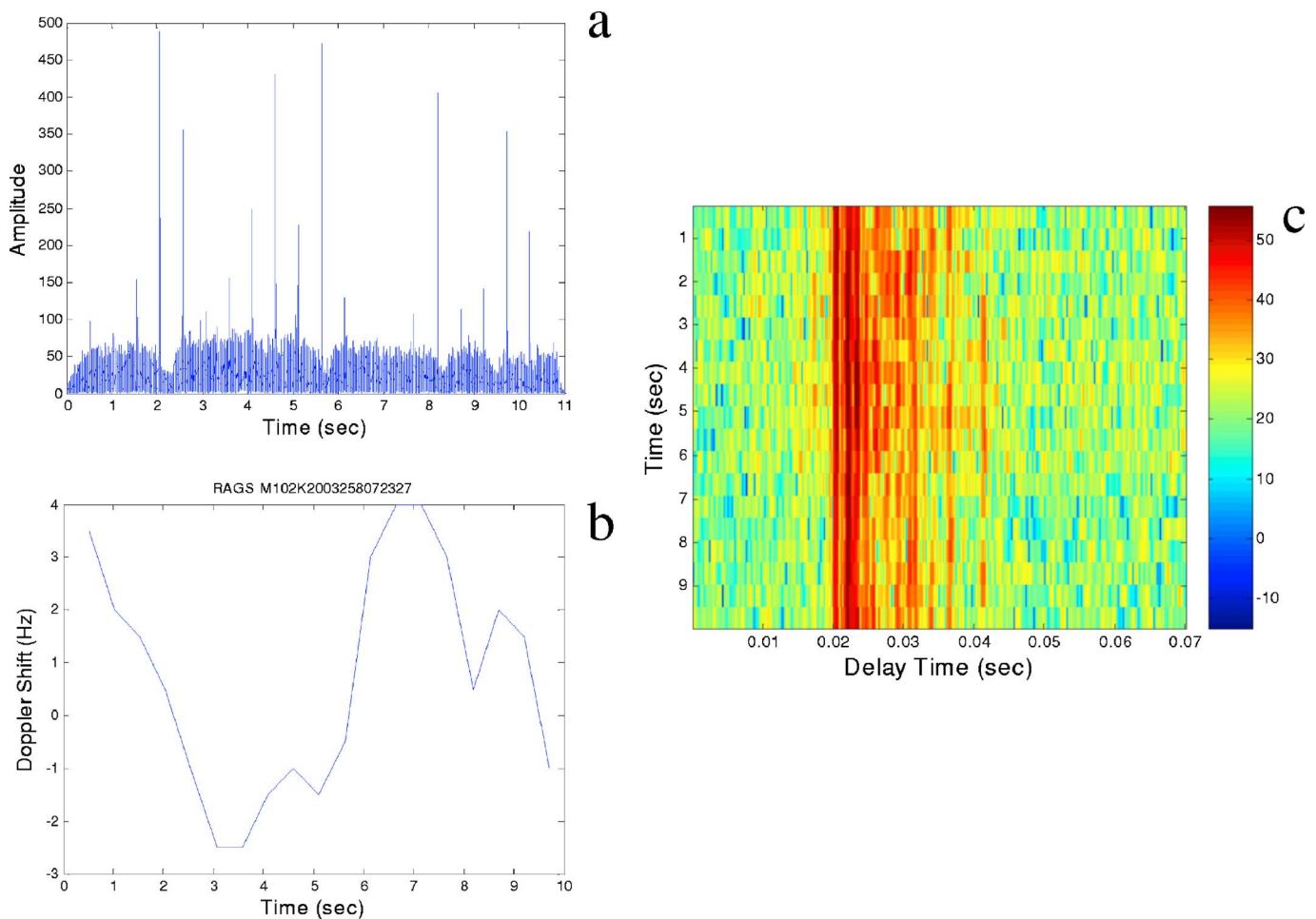
FIG. 19. (Color online) The effect of Doppler shift (array motion) on high frequency (20 kHz) signals. Matched filtered time series of the *m*-sequence signals without Doppler correction (a). Doppler shift measured from individual *m*-sequence signals using the wide band ambiguity function (b). Impulse response measured from the *m*-sequence signals after Doppler correction (c). Color scale is dB. Data were *m* sequences from the RAGS03 experiment. Source-receiver range is 3 km.

tion in the Doppler shift, which is nevertheless small compared with the carrier frequency. The small Doppler ratio (Doppler shift over the carrier frequency) indicates a small source/receiver motion which is consistent with the cable/array motion. After applying the Doppler correction, the measured channel impulse responses are plotted in Fig. 19(c). The amplitudes of the Doppler-corrected impulses vary slightly within the packet.

The above analysis shows that Doppler correction is critical for Doppler sensitive waveforms such as the *m* sequences, particularly at high (20 kHz) frequencies. Without the Doppler correction, one could underestimate the signal coherence-time. At low frequencies, Doppler correction due to array motion is usually negligible since the Doppler shift is often small compared with the frequency resolution.

To estimate the temporal coherence of the signal, the impulses responses of the *m* sequences were processed for each packet of *m*-sequence data, as described in Sec. III A. The average of all the temporal coherence curves is plotted in Fig. 20(a), from which one finds a mean coherence-time $\tau_{0.8} \approx 0.5$ s. Figure 20(b) shows the "instantaneous" coherence-time of individual packets as a function of time. Figure 20(c) shows the histogram of the coherence-time. One finds a mean of 0.5 s and a STD of 0.3 s.

VI. CHARACTERISTIC PROPERTIES OF THE SIGNAL coherence-time

In this section we investigate the range, frequency, and environmental dependence of the signal coherence-time of the data reported above. For that purpose, the data are characterized in terms of simple empirical fits. The result suggests certain power dependence but such dependence needs, without saying, further testing with more data.

A. The range dependence

Given the measurement of the signal coherence-time $\tau_{0.8}$ at two ranges from the ASCOT01 data, one can study the range dependence of the signal coherence-time. The measured (mean) signal coherence-time is plotted in Fig. 21 as a function of range for all three frequencies (450, 1200, and 3550 Hz) in a log-log plot. One finds that the data are well fitted with an inverse square-root power dependence on range [see Eqs. (6) and (7)]. Three lines are used to fit the data, which are given by $\tau_{0.8} = \delta R^{-1/2}$, where $\delta = 2.5 \times 10^3$, 6 $\times 10^3$, and 3.2×10^4 for 3550, 1200, and 450 Hz, respectively. The unit of τ is s and R is m.
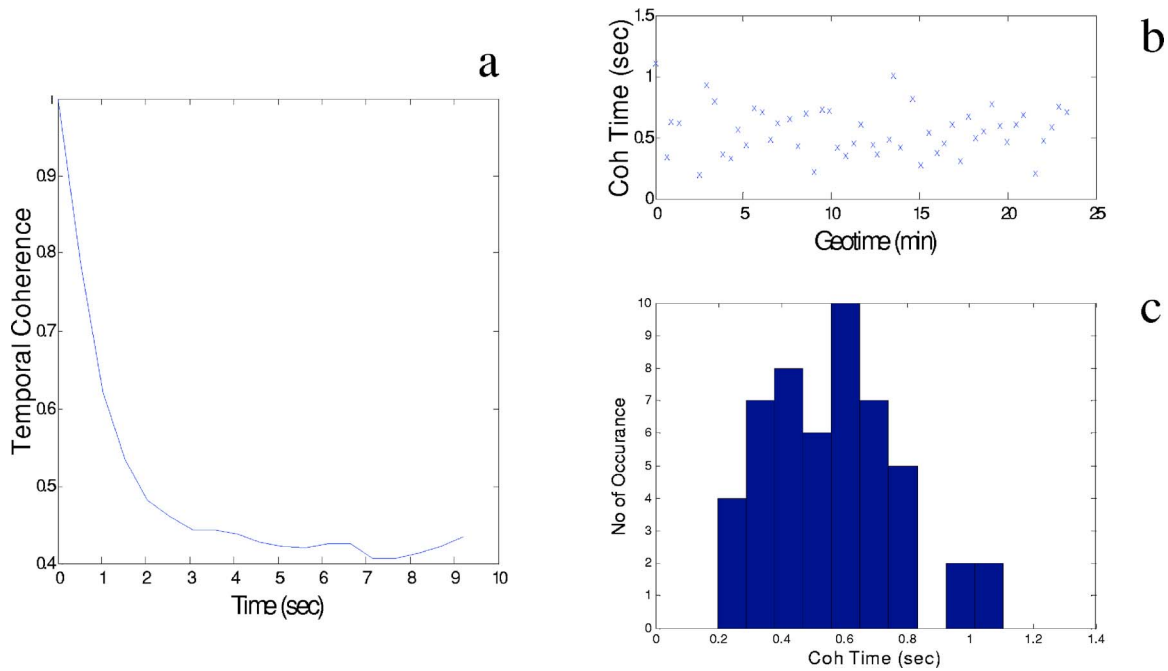
FIG. 20. (Color online) Temporal coherence of high frequency (20 kHz) signals from the RAGS03 experiment. Average temporal coherence of the *m*-sequence signals as a function of the signal separation time (a). "Instantaneous" signal coherence-time as a function of geotime (b). Histogram distribution of the signal coherence-time (c).

Deep water data showed that the coherence-time decreases as the inverse square root of the range at long ranges; more precisely, the phase rate increases as a square-root power of range.[25] This range dependence was derived theoretically using the Rytov approximation, as discussed above. At short ranges, the phase rate data in the deep ocean indicate a 3/2 power dependence on range for ranges less than 1 km.[25] The above data suggest that the signal coherence-time in shallow water decreases as the square root of range as in deep water. Whether the coherence-time in shallow water at short ranges (also) has a different range dependence is not known.

The ASCOT01 data also contained measurements at 2 km, which could, in principle, be used to address the range dependence. The data are not used in this paper because the data have a high variance and are not conclusive. Note that at a range of 2 km, the signal coherence varies significantly depending on whether the solitary internal waves are within the source-receiver range or not. (At long ranges, one can count on an average number of solitary internal waves in the range.) As such, data over an extended period of time are needed for the temporal coherence measurement to be statistically meaningful. The 2 km range data had a short duration;
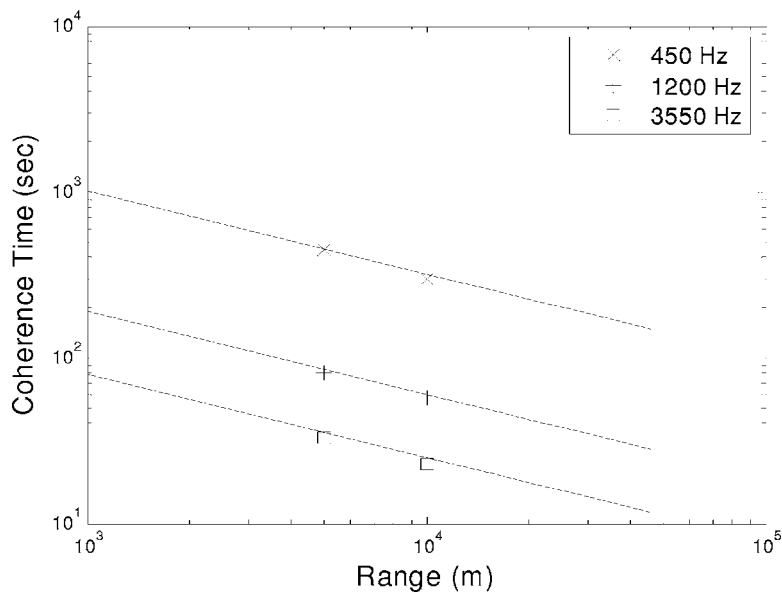
FIG. 21. Signal coherence-time plotted as a function of the range at three different frequencies, based on AS-COT01 data at 5 and 10 km ranges.

TABLE II. coherence-time $\tau_{0.8}$ and the spread of the coherence-time (in parentheses) measured off the coast of New Jersey; the later reflects the amount of the temporal variation of the coherence-time due to changes in the propagation environment.

Frequency	$\tau_{0.8}$ at 42 km (SWARM 95)	$\tau_{0.8}$ at 3 km (RAGS03)	$\tau_{0.8}$ at 10 km (Extrapolated)
20 kHz		0.5 (±0.3) s	0.27 (±0.16) sec
400 Hz	40 (±35) s		82 (±72) sec
224 Hz	130 s		266 sec

the short duration may be the reason for the large variance in the 2 km data compared with the mean coherence-time.

B. Frequency dependence

Based on the theoretical work in deep water [Eq. (7)], the coherence-time [Eq. (6)] is expected to decrease as inverse square-root power of range, and inverse linear power of frequency,

$$\tau_{0.8} = (0.11/\gamma)f^{-1}R^{-1/2}. \tag{11}$$

Assuming the inverse square-root dependence on the range (as studied above), one can study the frequency dependence of the coherence-time $\tau_{0.8}$.

The SWARM95 and RAGS03 data were collected at a different range than 10 km (Table II). Using the inverse square-root range dependence, one can extrapolate the coherence-time $\tau_{0.8}$ to a range of 10 km. The results are shown in Table II, last column. The 10 km data are plotted in Fig. 22. One finds that the data can be fitted with an inverse 3/2 power dependence on frequency, as shown by the curve labeled C.

The coherence-time $\tau_{0.8}$ measured from the ASCOT01 data at the 10 km range (Table I) is also plotted in Fig. 22. One finds that the data can also be fitted with an inverse 3/2 power dependence on frequency, as shown by the curve labeled B.
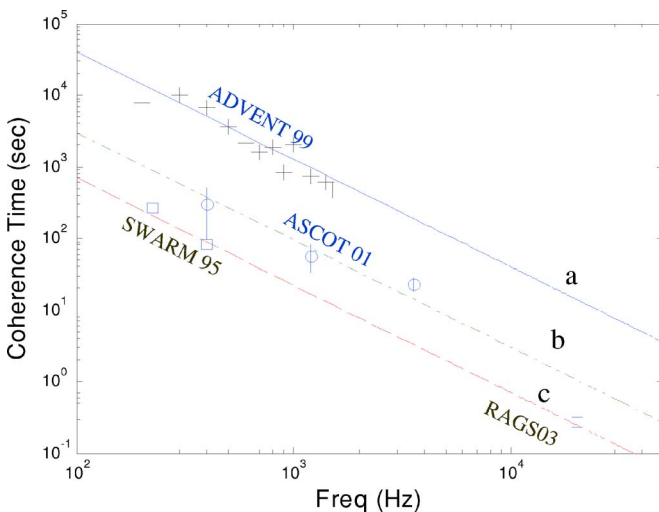
FIG. 22. (Color online) Signal coherence-time plotted as a function of signal frequency based on data from four different experiments. Dashed curves are fit to the data, as described in the text.

In Fig. 22 is also plotted the coherence-time $\tau_{0.8}$ measured during the ADVENT99 experiment, which was presented in Ref. 13. The temporal coherence was measured using matched field correlation, Eq. (8). The data were narrow band signals (200 to 1200 Hz in steps of 100 Hz) received on a vertical line array. The source receiver range was 10 km. One finds that the data are consistent with an inverse 3/2 power dependence on frequency, as shown by the curve labeled A.

In summary, one obtains an empirical fit of the data presented above as

$$\tau_{0.8} = \beta f^{-3/2}R^{-1/2}, \tag{12}$$

where $\beta = 4 \times 10^9$, 3×10^8, and 7×10^7 (m/s)$^{1/2}$ for the ADVENT99, ASCOT01, and SWARM95/RAG03 data, respectively. Using Eq. (6), one finds

$$\nu = \alpha f^{3/2}R^{1/2}, \tag{13}$$

where $\alpha = 0.11/\beta$. One thus finds $\alpha = 0.28 \times 10^{-10}$, 3.6×10^{-10}, and 1.6×10^{-9} (m/s)$^{-1/2}$ for the ADVENT99, ASCOT01, and SWARM95/RAG03 experiments, respectively. The 3/2 power dependence on frequency has not been predicted before.

C. Relation to the sound speed STD

The more random the ocean becomes, the less the signal coherence-time will be. Assuming that the medium fluctuation is the cause for the signal coherence loss, the question is how is the coherence loss related to, or more specifically, how is β (or α) related to the oceanographic activities in the water column. This topic obviously requires a detailed modeling analysis and could benefit from some theoretical predictions. For the purpose of data characterization, it is useful to determine the relation of the above coefficient to the sound speed STD, noting that the latter is related to the amplitude of the internal waves and other fine structures.

Let us examine the sound speed STD in the various experiments. The four experiments, ADVENT99, ASCOT01, SWARM95, and RAGS, all took place in a summer environment with a layer of warm water over the cold water. The water depth was approximately 81 m for the ADVENT99 experiment, 100 m for the ASCOT01 experiment, and varied from 57 to 88 m in the SWARM95 experiment. The average sound speed profiles and the sound speed STD can be found in Refs. 13, 22, and 26 for the ADVENT99, ASCOT01, and SWARM95 environments, respectively. The RAGS experiment took place in the same area as SWARM95 and has a similar sound speed profile as SWARM 95. The sound speed is nearly constant below 30 m with sound speed increasing slightly (ADVENT99), gradually (ASCOT01), or sharply (SWARM95) to a high sound speed found in the warm layer. The gradual to sharp transitions between the warm and cold layers are responsible for different magnitudes of internal waves, which reside primarily at the pycnocline, which often coincides with the thermocline. The primary difference between the three environments is the sound speed STD, which varies from 0.3 m/s in ADVENT99, to 2–3 m/s in ASCOT01 and 5 m/s in SWARM95, as measured at the depth

TABLE III. Environmental dependence.

	Water depth (at receiver)	Thermocline depth	Sound Speed STD (δC)	Alpha of Eq. (13)
ADVENT99	80 m	~5 m	0.3–0.6 m/s	0.027×10^{-9}
ASCOT01	100 m	~8 m	2–3 m/s	0.28×10^{-9}
SWARM95	88 m	~20 m	5 m/s	1.74×10^{-9}

of the thermocline. (The thermocline is not well defined in the ADVENT99 experiment. Because of the uncertainty of sound speed measurement within 5 m of the surface, the data at 10 m are used for ADVENT99.) The acoustic parameters are summarized in Table III.

The sound speed STD (at the pycnocline) is a good indicator of the strength of the internal waves as it is proportional to the amplitude of the internal waves to the first order of approximation.[19] The Adventure Bank has little internal wave activity. The coasts of New England and New Jersey have medium to high amplitude internal waves. Internal waves create shear instabilities and produce turbulence at the boundaries. These instabilities can affect high frequency sound propagation.

To determine if and how the medium fluctuations affect the signal coherence-time (at a fixed frequency and range), one needs to determine the correlation between the sound speed STD (at the thermocline) and the coefficient α of the signal phase rate, Eq. (13). We plot in Fig. 23 the value of α as a function of the sound speed STD as given in Table III. The α data are consistent with a quadratic power dependence on the sound speed STD as shown by the dashed curve; $\alpha = 0.6 \times 10^{-10} (\delta C)^2$.

D. Observations

The above analysis provides an empirical parameterization of the signal coherence-time with respect to the signal frequency, source receiver range, and sound-speed STD. The result is useful for system design and analysis.

From the physics point of view, the question is how is the signal coherence loss related to the environmental conditions—how to predict the signal coherence loss. As remarked above, there exists (to our knowledge) no theoretical analysis for the signal coherence loss in a time varying shallow water environment. The problem is notably difficult, and is beyond the scope of this paper. Nevertheless, some observations can be made from the data and will be discussed later.

It is generally accepted that the signal coherence-time reflects the time scale of the processes responsible for the signal coherence loss. As such, one can infer from the measured coherence-time the underlying processes causing the signal coherence loss. Signal coherence loss can be caused by signal scattering from the medium heterogeneities, rough surfaces, and bottom inhomogeneities. Which processes are dominant depends on the signal frequency, the magnitude of the inhomogeneities, and the roughness of the boundaries. To properly interpret the data, one needs to model the signal propagation that depends on the sound speed profile and bottom attenuation. The measured channel impulse response should be used to guide and verify the modeling results.

Normally, acoustic rays with shallow depletion/elevation (D/E) angles are trapped in the sound waveguide and thus are most sensitive to medium inhomogeneities such as internal waves and turbulence. On the other hand, acoustic rays with steep D/E angles tend to interact more with surfaces and bottom and are thus more affected by the surface roughness, bottom attenuation, and changes in the overall sound speed. To study the effect of the rough surfaces in a downward refractive sound speed profile, it is preferred to use a source near the surface.

The ASCOT01 experiment used a source near the bottom. As noted above, the late arrivals in the ASCOT01 data (e.g., the 3550 and 1200 Hz signals) change slowly with time. The change in arrival time is, in this case, caused by the change of the layer depth (due to internal tide) rather than the surface height (the external tide) or roughness (wind speed). The reason, as pointed out in Ref. 23, is that the high
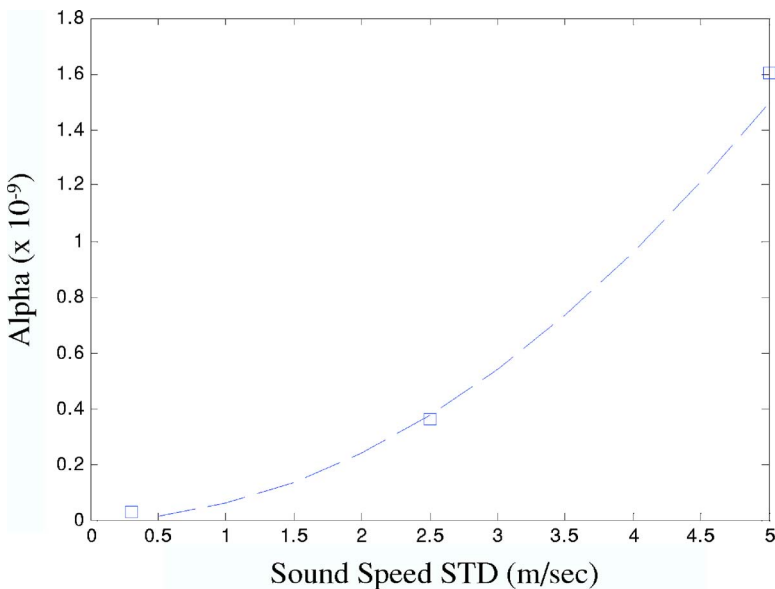
FIG. 23. (Color online) The coefficient α measured from the data (square) as a function of the sound speed STD.

D/E angle rays have been attenuated, so the signal consists mostly of refracted-bottom-bounced rays. This has also been demonstrated with modeling results in other shallow water environments.[27,28] One finds that the signals from a bottom mounted source, propagated to long ranges, have generally little interaction with the surface in a downward refractive sound speed profile.

Despite the evidence that internal tides affect the multipath arrival time at high frequencies, the internal tides have negligible effect on the signal coherence-time at 3550 Hz in the ASCOT01 environment. This is due to the fact that the signal coherence-time at high frequencies (at the ASCOT01 site) is short (of the order of tens of seconds). This time scale is indicative of the microfine structures (e.g., turbulence). Figure 1 shows rapid fluctuations of the multipath arrivals with time scales on the order of minutes. The results are consistent with the observation that turbulence (and internal waves) has a significant impact on high frequency sound propagation. The instantaneous coherence-time at 3550 Hz [Fig. 4(b)] shows a pattern of random fluctuation, which is expected from random structures like turbulence.

The low frequency (450 Hz) data from the ASCOT0l experiment show, on the other hand, a very stable arrival pattern, particularly for the early arrivals (Fig. 8). (The impulse responses at time 250–270 min are out of alignment, an artifact of data processing.) The instantaneous coherence-times [Fig. 9(b)] are of the same order of magnitude except for certain periods (between 50–100 min in Fig. 8), where the fifth arrival has a high amplitude. Similar observations can be made in Fig. 14 when the third arrival has a high amplitude. At these time periods, the instantaneous coherence-time is higher than other time periods. This is consistent with the fact that a higher coherence-time is obtained when the signal energy is concentrated in a small number of stable multipaths. It is presumed that during these periods, the signals have less scattering and thus have high amplitudes.

The instantaneous coherence-time at 1200 Hz [Fig. 7(b)] shows periodicities on the order of tens of minutes. Similar periodicities are also observed at 450 Hz [Fig. 9(b)]. The origin of this is not known.

The 400 Hz signal from the SWARM95 experiment shows very different properties than the 450 Hz signal from the ASCOT01 experiment, despite the fact that they propagate under similar sound speed profiles. The multipath structures are not as clearly defined in Fig. 16 as in Fig. 8. The multipath arrival times are not as stable in Fig. 16 as in Fig. 8. The reason is that the internal waves off the coast of New Jersey are much stronger than that off New England (See Sec. VI C). Internal waves induce mode couplings in the SWARM95 data,[14,24,27] which smear the arrival pattern as well as temporal stability of the received signals. One notes also that for the 450 Hz signal from the SWARM95 data, the arrival time difference between the early and late arrivals remain approximately unchanged over an 18 h period (Fig. 16), whereas the arrival time difference changes with the tidal cycle in the ASCOT01 data (as discussed above for the 3550 and 1200 Hz signal, a similar pattern was also seen for 450 Hz signal in Fig. 8). The effect of the internal tides on the signal arrival time is also expected for the SWAM95 signal, but such arrival time differences are easily diminished by the mode coupling effect.

The 224 Hz signal in the SWARM95 data show a dominant group of arrivals with approximately the same total delay time as that seen in the 400 Hz signal. Individual multipath arrivals are not identified due to the narrow band (16 Hz) of the signal. This is not an environmental effect.

VII. SUMMARY

In this paper, we conducted a systematic measurement of the temporal coherence of acoustic signals propagating in three shallow water areas. The three environments represent areas of low, medium, and high internal wave activity. The measurements cover low (below 1.5 kHz), mid (2 to 5 kHz), and high frequency data. The receiver range covers 3, 5, 10, and 42 km. The "instantaneous" temporal coherence of the signals is also measured, from which one obtained the time variation of the signal coherence-time as well as its mean and STD. To achieve a statistical mean, the temporal coherence is averaged over hours of data, which are generally considered adequate for high frequency signal fluctuations. For low frequency acoustic signals in an area where solitary internal waves are dominant, data beyond 12 h were used to include the solitary internal wave effect. Otherwise, hours of data should be adequate (for diffuse internal waves). An empirical analysis is conducted on the relationship between the signal coherence-time and the acoustic frequency, the source receiver range, and the sound speed STD. It is found that the dependence on frequency and sound speed STD deviate significantly from a previous theoretical prediction for deep water.

[1] R. C. Spindel, R. P. Porter, and R. J. Jaffee, "Long-range sound fluctuations with drifting hydrophones," J. Acoust. Soc. Am. **56**, 440–446 (1974).

[2] J. G. Clark and M. Kronengold, "Long period fluctuations of CW signals in deep and shallow water," J. Acoust. Soc. Am. **56**, 1071–1083 (1974).

[3] R. E. Williams, "Creating an acoustic synthetic aperture in the ocean," J. Acoust. Soc. Am. **60**, 60–73 (1976).

[4] R. E. Williams and H. F. Battestin, "Time coherence of acoustic signals transmitted over resolved paths in the deep ocean," J. Acoust. Soc. Am. **59**, 312–328 (1976).

[5] T. M. Georges, L. R. Boden, and D. R. Palmer, "Features of the Heard Island signals received at Ascension," J. Acoust. Soc. Am. **96**, 2441–2447 (1994).

[6] T. G. Birdsall, K. Metzger, M. A. Dzieciuch, and J. Spiesberger, "Integrated autocorrelation phase at one period lag," J. Acoust. Soc. Am. **96**,

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Tsih C. Yang: Temporal coherence of acoustic signals   2613

2353–2356 (1994).

[7]R. M. Fitzgerald, A. N. Guthrie, and J. D. Shaffer, "Low-frequency coherence transverse to the direction of propagation," J. Acoust. Soc. Am. **60**, 752–753 (1976).

[8]I. Dyer, "Ocean dynamics and acoustic fluctuations in the Fram Strait marginal ice zone," Science **236**, 435–436 (1987).

[9]W. H. Munk and F. Zachariasen, "Sound propagation through a fluctuating stratified ocean: Theory and observation," J. Acoust. Soc. Am. **59**, 818–838 (1976).

[10]S. M. Flatte *et al.*, *Sound Transmission Through a Fluctuating Ocean*, edited by S. M. Flatte (Cambridge University Press, Cambridge, 1979).

[11]P. L. Nielsen, M. Siderius, and F. B. Jensen, "Acoustic time-variability measurements in the strait of Sicily," *Proc. of the 5th European Conference on Underwater Acoustics*, Lyon, France, 11–14 July, 2000, pp. 57–62.

[12]T. C. Yang and M. Siderius, "Temporal coherence and fluctuation of acoustic signals in shallow water," In Ref. 11, pp. 63–68.

[13]T. C. Yang, K. Yoo, and M. Siderius, "Internal waves and its effect on signal propagation in the Adventure Bank," *Proc. of 8th International Congress On Sound And Vibration*, Hong Kong, 2-6 July 2001, pp. 3001–3008.

[14]R. H. Headricket *et al.*, "Acoustic normal mode fluctuation statistics in the 1995 SWARM internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 201–220 (2000).

[15]D. Rouseff *et al.*, "Coherence of acoustic modes propagating through shallow water internal waves," J. Acoust. Soc. Am. **111**, 1655–1666 (2002).

[16]P. C. Mignerey and M. H. Orr, "Observations of matched field autocorrelation time in the South China Sea," IEEE J. Ocean. Eng. **29**, 1280–1291 (2004).

[17]M. Badiey, J. Simmen, and S. Forsythe, "Frequency dependence of broadband propagation in coastal regions," J. Acoust. Soc. Am. **101**, 3361–3370 (1997).

[18]Preliminary results were presented in T. C. Yang, "Environmental effects on phase coherent underwater acoustic communications: A perspective from several experimental measurements," *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (American Institute of Physics, Melville, NY, 2004), pp. 90–97.

[19]F. Dyson, W. Munk, and B. Zetler, "Interpretation of multipath scintillations Eleuthera to Bermuda in terms of internal waves and tides," J. Acoust. Soc. Am. **59**, 1121–1133 (1976).

[20]P. H. Dahl, a. B. Baggeroer, P. Mikhalesky, and I. Dyer, "Measurement of the temporal fluctuations of CW tones propagated in the marginal ice zone," J. Acoust. Soc. Am. **83**, 2175–2179 (1988).

[21]J. Sellschopp, P. Nielsen, and M. Siderius, "Combination of acoustics with high-resolution oceanography," *Impact of Littoral Environmental Variability on Acoustic Prediction and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, The Netherlands, 2002), pp. 19–26.

[22]J. Sellschopp, "High resolution measurements of the ocean fine structure and their relation to sound transmission," *Proc. of the Internal Conference "Underwater acoustic measurements: Technologies and Results,"* Crete, Greece, June 2005.

[23]P. Nielsen, M. Siderius, and J. Sellschopp, "Broadband acoustic signal variability in two typical shallow-water regions," in Ref. 21, pp. 237–244.

[24]J. R. Apel *et al.*, "An overview of the 1995 swarm shallow water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–500 (1997).

[25]K. D. Rolt and P. A. Abbot, "Littoral coherence limitations on acoustic arrays," *Acoustic Imaging*, edited by S. Lees and L. A. Ferrari (Plenum, New York, 1997), Vol. **23**.

[26]K. Yoo and T. C. Yang, "Broadband source localization in shallow water in the presence of internal waves," J. Acoust. Soc. Am. **106**, 3255–3269 (1999).

[27]S. Finette *et al.*, "Acoustic field variability induced by time evolving internal wave fields," J. Acoust. Soc. Am. **108**, 957–972 (2000).

[28]M. B. Porter, S. M. Jesus, Y. Stephan, X. Demoulin, and E. Coelho, "Tidal effects on source inversion," in *Experimental Acoustic Inversion Methods for Exploration of the Shallow Water Environment* edited by A. Caiti, J.-P. Hermand, S. M. Jesus, and M. B. Porter (Kluwer Academic, New York, 2000).

High-frequency channel characterization for M-ary frequency-shift-keying underwater acoustic communications

Wen-Bin Yang[a] and T. C. Yang
Naval Research Laboratory, Washington D.C. 20375

Many ($>$100) packets of pseudo-random signals transmitted during the TREX04 experiments are analyzed in this paper to deduce the narrowband signal envelope amplitude statistics over a wide band (15–19 kHz) of frequencies. The envelope amplitude statistics are found to be non-Rayleigh-type with a long-tail distribution at high amplitudes. Long-term and short-term fading statistics are deduced from the data, exhibiting near lognormal and Rayleigh distributions, respectively. The combined distribution is closely approximated by a K-distribution, which fits the measured amplitude distributions well over a wide band of frequencies. [DOI: 10.1121/1.2346133]

I. INTRODUCTION

As is well known, the communication bit error rate (BER) is heavily influenced by the channel characteristics. Underwater acoustic communication channels differ from radio frequency (RF) communications channels by: (i) A longer multipath delay, which extends over many symbols causing intersymbol interference (ISI), (ii) a higher Doppler spread which implies a shorter channel coherence time, and (iii) a time-varying Doppler shift due to the relative platform speed compared with the sound speed. Under the assumption of wide sense stationary and uncorrelated scattering, the channel properties can be characterized by the scattering function as a function of delay time and Doppler frequencies.[1] The scattering function becomes the channel impulse responses for stationary sources and receivers. The latter is often used to assess the effect of the medium on communication performance. For time-varying oceans, channel characterization is concerned with a statistical description of the impulse response functions (the amplitudes of individual arrivals), which is a study in the time domain. Channel characterization can also be done in the frequency domain using the channel transfer functions (the Fourier transform of the channel impulse responses). Although the two are, in theory, equivalent, they have different implications in practice. This paper addresses the channel characterization in the frequency domain, specifically the signal amplitude envelope statistics, which forms the basis of BER predictions for M-ary frequency-shift-keying (MFSK) signals. The time domain channel characterizations form the basis for performance predications for M-ary phase-shift-keying (MPSK) signals. While the multipath amplitude fluctuation analysis is outside the scope of this paper, the physical insights on multipath fluctuations are needed to understand/interpret the frequency domain results.

Figure 1 shows a snapshot of high-frequency channel impulse responses measured from at-sea data over a period of a little more than 10 s. The signal has a center frequency of 17 kHz and a bandwidth of 5 kHz. Details are given later in the paper. One notes that the channel impulse responses vary on a scale of tenths of a second. The channel coherence time is on the order of 0.2 s. In Fig. 1, one observes five to six strong multipath arrivals, plus numerous late arrivals. Not only does the amplitude of each path change with time, the arrival times of individual paths may also vary with time. This property is different from impulses responses at low ($<$5 kHz) frequencies, where the signal has a coherence time of tens of second and the channel impulse responses remain coherent within a communication packet.[2] Channel characterization (in the time domain) requires a determination of the amplitude and phase (fluctuation) statistics for individual path arrivals. Galvin and Wang[3] measured the fluctuation statistics for a direct path and a surface reflected arrival using parametric sonar. The acoustic signal is centered at 5 kHz with a 2 kHz bandwidth. They found that the amplitudes follow a Rician distribution and the phase statistics is Gaussian-type. There are very few direct experimental measurements in the published literature. The difficulties are that the multipaths arrivals are often clustered together and not easily separable except at short ranges.

The above experimental findings serve as the basis for certain channel characterization models. Geng and Zelinski[4] assumed that the received signal is comprised of a dominant stable component and many smaller, randomly scatter components; hence, the signal follows the Rician fading statistics. Estes *et al.*[5] proposed a different model in which a raised cosine windowing function is used to introduce the nonstationary characteristics of the channel due to platform and boundary movement. Jobst *et al.*[6] proposed a stochastic model in which the temporal and spectral fluctuations are not Rayleigh distributed. The number of eigenrays reaching the receiver is Poisson distributed, with a mean value determined by the Ray theory. Galvin and Coates[7] model the channel using a combination of linear and nonlinear transformation on a Gaussian variables. One notes that there are many channel models in literature. The reasons are due to a lack of: (1) Theoretical understanding of the path fluctuations and the corresponding statistics, and (2) direct experimental measurements of the amplitude fluctuations of individual paths.

---
[a]Electronic mail: wyang@wave.nrl.navy.mil
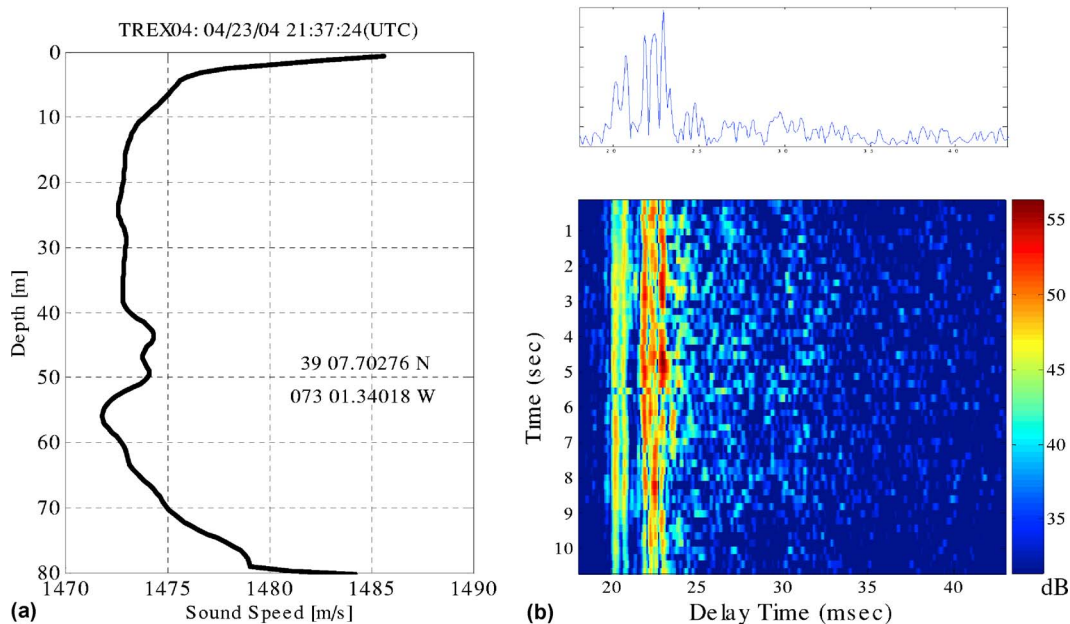
FIG. 1. (Color online) (a) A representative sound speed profile collected during the TREX04 experiment. (b) Channel impulses responses measured from one packet of M-sequence data.

Note that the fluctuation statistics are path and frequency dependent. At low frequencies, one observes that the early arrivals are relatively stable, whereas the late arrivals are random and varying fast with time. The reason is that early arrivals are affected by medium inhomogeneities, which at low frequencies are caused by internal waves, and the late arrivals are scattered returns from the surface and bottom roughness. Since the early arrivals are more dominant, the coherence time is of the order of tens of seconds to minutes (the scale of internal waves). This is not the case at high frequencies, as evidenced by the much shorter coherence time. At high frequencies, the microfine inhomogeneities, e.g., the turbulence, play an important role. A proper channel model must incorporate the physics of signal fluctuations. Whether theoretically motivated or experimental-data based, one needs a channel model to predict the BER for MPSK signals.

For MFSK signals, the symbols are spread over the frequency band and modulated in both frequency and time. Channel characterization is now concerned with the envelope amplitude fluctuation statistics as a function of frequency, i.e., the fluctuations statistics of the amplitudes of the channel transfer function for a given frequency. This is called the narrowband fluctuation statistics, since the frequency bin size is much smaller than the center frequency of the whole band. (Only the absolute value of the amplitude is of interest here, since each symbol is detected by an energy detector.) The BER for MFSK signals depends on the envelope amplitude fading statistics as a function of frequency. Rayleigh and Rician amplitude probability distributions are two commonly assumed models for signal fading in RF communications.[8–10] Narrowband signal fluctuation statistics have been studied at low frequencies (e.g., <1 kHz) based on the theory of sound propagation through internal waves.[11] Rayleigh and Rician statistics are associated with saturated and

partially saturated schemes, in which the multipaths are totally or partially random.[12] For acoustic communications, Catipovic et al.[13] assumed a Rayleigh distribution for MFSK signals.

The envelope amplitude fluctuation statistics for a fixed frequency are very different from that for a fixed path. As remarked above, the amplitudes of some paths may follow a Rician distribution, whereas the other may follow a Rayleigh distribution or something else. In the frequency domain, all multipath arrivals contribute to the fluctuations. Hence, a priori, what the envelope amplitude statistics will be is not clear. (If the time domain amplitude distribution is Rayleigh, the frequency domain amplitude distribution will also be Rayleigh. But this is not the case for the majority of the underwater acoustic channels.) To our knowledge, there are no direct experimental measurements of the envelope amplitude statistics for high-frequency MFSK signaling.

The MFSK symbol duration is, normally by design, of the order of the multipath spread. To avoid ISI interference, the symbol duration (including the guard time if appropriate) should be longer than the multipath spread; but, in practice, this is often not the case. In that case, there will be some ISI coming from the late arrivals beyond the MFSK symbol duration. For MFSK performance prediction, the amplitude statistics used should include the ISI interference according to the signal symbol duration (or the frequency bin size) selected.

For MFSK symbol duration on the order of the multipath delay, one can assume that the (narrowband) signals between adjacent frequency bins are uncorrelated. More specifically, the frequency bin size needs to be greater than the frequency coherence bandwidth, which is of the order of the inverse of the multipath spread. In this case, the envelope amplitude statistics between the different frequency components can be assumed to be independent and identically dis-

tributed, as the channel property is expected to vary slowly with frequency. The data analysis presented below supports this hypothesis.

In this paper, narrowband amplitude envelope statistics are deduced from data collected at sea. The measured distributions deviate significantly from the Rician and Rayleigh distributions and are found to be close to the K-distribution. A physics-based interpretation of the K-distribution is provided. The implication is that the BER for MFSK underwater acoustic communications is worse than that in a RF channel for a given signal-to-noise ratio (SNR), since the BER for a K-distribution is worse than that of Rayleigh and Rician distributions. This is independently verified using frequency-hopping frequency-shift-keying data measured at sea.[14]

The paper is organized as follows. Section II describes the experimental data analysis. Narrowband envelope amplitude distribution statistics are deduced from data covering a wide band (4 kHz) of frequencies. Section III reviews the three amplitude statistical distribution models and compares them with data. Section IV presents the long-term and short-term fading statistics, from which a distribution close to the K-distribution is derived. Section V provides conclusions.

II. CHARACTERISTICS OF THE AMPLITUDE FLUCTUATIONS

A. MFSK modulation

For a narrowband signal, the channel transfer function can be defined by

$$R(t,f) = H(t,f)S(f), \qquad (1)$$

where $H(t,f)$ is the channel transfer function at frequency f, assumed to be time dependent, R is the received signal, and S is the source amplitude. MFSK signals consist of many narrowband signals at frequencies f_k, separated by Δf, where $k = 1, \ldots, K$,

$$R(t,f_k) = H(t,f_k)S(t,f_k), \qquad (2)$$

where $S(t,f_k)$ is the transmitted symbol sequence in frequency bin f_k at time t, and $S(t,f_k) = 0$ or 1. Each symbol has a time duration $\Delta t = 1/\Delta f$. Detection of symbols at the receiver is based on the symbol intensity $|R(t,f_k)|^2$, which is heavily influenced by the channel spectral level $|H(t,f_k)|^2$. Hence, BER modeling/prediction requires knowledge of the statistical distribution of $|H(t,f_k)|$, the envelope amplitude statistics.

To measure the envelope amplitude statistics, one can use

$$|H(t_m,f_k)|^2 = |R(t_m,f_k)|^2/|S(t_m,f_k)|^2, \qquad (3)$$

where $|R(t,f_k)|^2$ and $|S(t,f_k)|^2$ are the spectral level of the received and transmitted signals as a function of frequency at $t_m = m\Delta t$, for the mth symbol. Equation (3) provides a snapshot of the channel transfer function at a particular time as a function of frequency. In the experiment described next, pseudo-random noise signals are used. The pseudo-random signal has a flat spectrum when averaged over many sampling periods. Using $\Delta t = 12.5$ ms, more than 10^5 numbers of realizations of Eq. (3) are obtained from data

to determine the amplitude envelope statistics at a fixed frequency.

The alternative is to use very short pulses or single-frequency continuous waves. For a pulse, if the source has a flat spectrum, i.e., $S(t_m,f_k) \approx S_0$, one has

$$|H(t_m,f_k)|^2 = |R(t_m,f_k)/S(t_m,f_k)|^2 \approx |R(t_m,f_k)|^2/|S_0|^2. \qquad (4)$$

For continuous waves, the source spectrum can be assumed to be a constant. Although in both cases, data processing is simplified somewhat, the shortcoming is that it will take tens of hours or more data to achieve the same statistics (total number of realizations) as obtained using the pseudo-random signals (based on the time-bandwidth product calculation).

B. TREX04 experiment

TREX04 experiment was conducted by the Naval Research Laboratory (Washington D.C.) in April 2004, which took place in the coast of New Jersey. Figure 1(a) shows a sound speed profile based upon measurement at the site. Acoustic communication data were transmitted from a fixed source to a fixed receiver array at the range of 3.4 km. Water depth in the experimental area is about 70 m. The source and receivers were located at about 35 m depth. The vertical array has an aperture of approximately 2 m, and contains eight hydrophones with nonuniform spacing. The data presented below are from a single receiver; little difference is found between the receivers. The data have a high SNR $\geqslant 30$ dB. The recording has a dynamic range 96 dB, so that the recorded data provide sensitive measurements of signal fading.

Various communication signals centered at 17 kHz were transmitted and received at an eight-hydrophone line array. The M-sequence signal was one of the communication signals collected in the TREX04 experiment. It uses a bandwidth of 5 kHz. Each transmitted packet lasted approximately 10.7 s, and contained 53 M-sequences. A total of 134 packets, extended over a period of 1 h and containing 7102 M-sequences, were analyzed.

Figure 1(b) shows the channel impulse responses measured from 53 M-sequences within one packet. The dominant arrivals are confined within 5 ms of arrival time, with numerous weaker late arrivals. Modeling results show these are refractive arrivals as steep; bottom interacting arrivals have been severely attenuated at a range of 3 km. As remarked above, at high frequencies, the signal fluctuates rapidly with geotime.

The M-sequence signals are used to measure the channel transfer function as a function of frequency and time. The M-sequence signals are often referred to as pseudo-random-noise signals, since they have an average spectrum similar to that of white noise. The source spectrum for 53 M-sequences is plotted in Fig. 2(a), where the signal is processed with a frequency bin size $\Delta f = 80$ Hz. The received M-sequence signals have a very different spectrum than the transmitted one. Figures 2(b)–2(d) show three received spectra as a function of frequency and time, which were transmitted at three separate times. One observes significant variations in the signal levels compared with the transmitted signal and with each other. This is a result of random media propagation, which
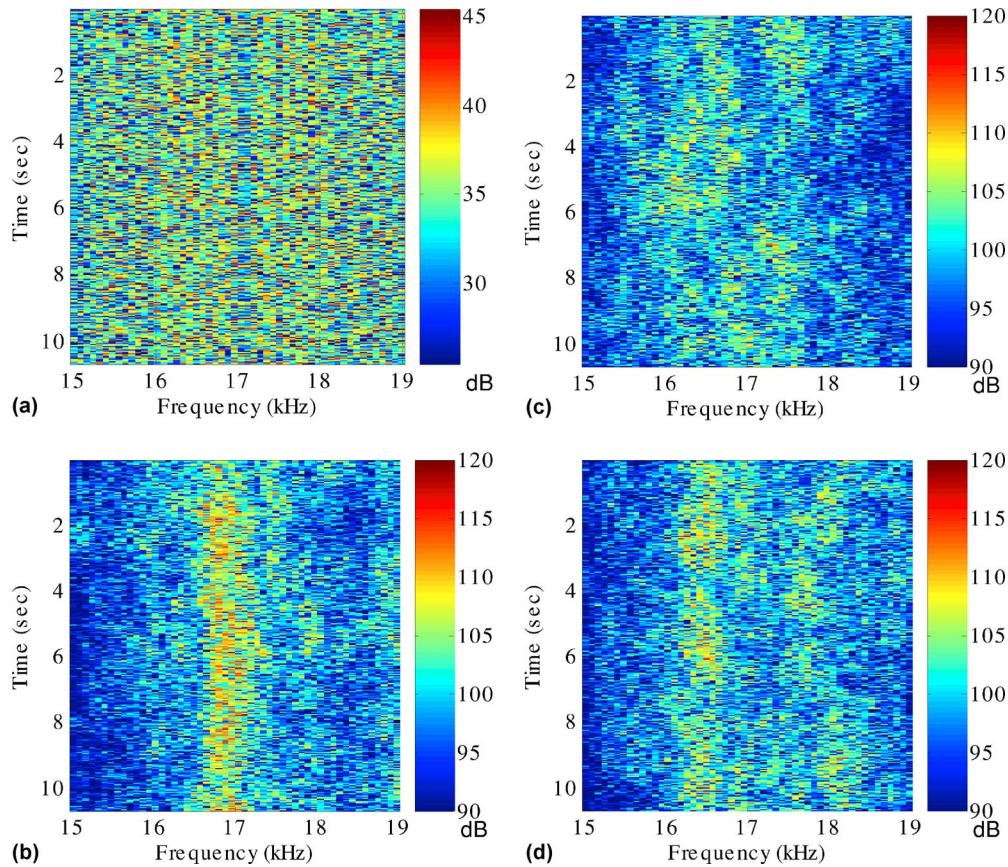
FIG. 2. (Color online) (a) The source spectrum: The spectrum of the M-sequence signal as a function of frequency and time. (b)–(d) The received M-sequence spectra as a function of frequency and time, based on three different packets transmitted at three different times. The frequency bin size is 80 Hz for all figures.

produces time-varying phases between modes or rays. Destructive interferences between modes or rays cause signal fading at some frequencies (frequency-selective fading). Note that the fading pattern is time dependent and varies significantly from packet to packet. Signal fading is a dominant source for BER for MFSK signals. The objective of the channel characterization in this paper is to determine the envelope amplitude statistics resulting from random media signal propagation.

For $\Delta f = 80$ Hz, the symbol duration is 12.5 ms. Compared with the impulse response shown in Fig. 1(b), one notes that a small portion of the current symbol will leak to the next symbol due to multipath arrivals beyond 12.5 ms, causing a small ISI. Note that for $\Delta f = 160$ Hz, there will be more ISI. On the other hand, for $\Delta f = 40$ Hz, the symbols are ISI free. The ISI is part of the amplitude envelope statistics that must be included for BER calculation/prediction. Thus, in theory, the amplitude envelope distribution will be dependent on the frequency bin size. The data analysis results shown next show a small frequency bin size dependence.

A word of caution is warranted here. One notes that the frequency components (bins) are, in theory, orthogonal to each other; implying that there is no leakage of the symbol energy from one frequency channel to the other. In practice, this may not be the case if there is significant source/receiver motion or rapid time variation in the propagation channel; causing frequency shift for the MFSK symbols. An error in the Doppler shift estimation could cause an energy leak from

adjacent frequency bins (referred to as the co-channel interference). To minimize this effect, the frequency bin width, Δf, is often chosen to be much larger than the uncertainty in the Doppler shift estimation. In this paper, data from fixed sources to fixed receivers are used with near-zero Doppler shift. For this data, the co-channel interference is expected to be negligible.

C. Envelope amplitude statistics

The M-sequence data were transmitted in consecutive packets; each of 15 s duration with built in time gaps. Total transmission time was ~40 min. The M-sequence data are processed first by removing the transducer's frequency response from the received data. The beginning of the M-sequence in each packet is determined by matched filtering the data using either the probe signal before the M-sequence or the first M-sequence; a standard processing technique for communications. The M-sequence data are then Fourier transformed with a window size equal to the symbol duration. The channel spectrum is obtained using Eq. (3); with the received and transmitted data processed in the same way.

One obtains the mean spectral level as a function of frequency, $\Sigma(f_k)$, by summing the spectrum level over all data samples and dividing the result by the number of samples. It is found that $\Sigma(f_k)$ decreases by as much as 5 dB at the edges of the frequency band. Since Eq. (3) is unstable
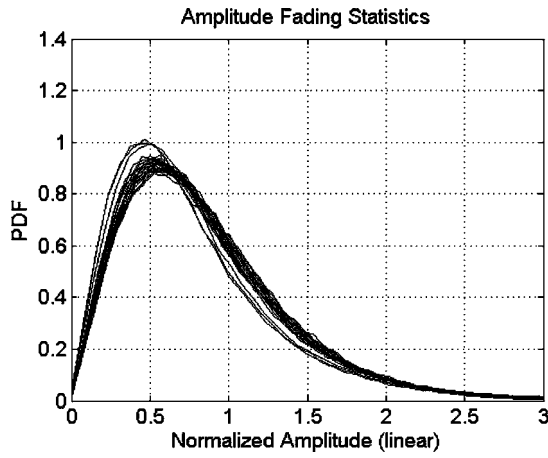
FIG. 3. Amplitude fading statistics for individual frequency bins (frequency bin size=80 Hz).

close to the edges of the frequency band (where the source spectrum is near zero), the edge frequencies are discarded in our analysis. $\Sigma(f_k)$ in the middle frequency band ($\sim$4 kHz), representing the average system response, is expected to vary little within the frequency band (which is much smaller than the center frequency). The measured $\Sigma(f_k)$ varies by 1–2 dB, and is attributed to the uncertainty in the transducer response curve, which was undersampled in the original (calibration) measurements. This 1–2 dB variation is removed by the following operation:

$$|H'(f_m,f_k)|_{\text{new}} = |H(t_m,f_k)|/\sqrt{\Sigma(f_k)}, \quad (5)$$

where t_m denotes the symbol sequence in time. Henceforth, the corrected data will be denoted by $|H(t_m,f_k)|$ without the "prime."

As the data were transmitted in packets, one can represent the realization in time as $t_m=nT_0+j\Delta t$, where n is the packet number, $n=1,\ldots,134$, T_0 is the time separation between packets, and j is the symbol number within a packet, $j=1,\ldots,856$ and $\Delta t=1/80$ s for $\Delta f=80$ Hz. This yields a total of 114 704 realizations in time for each fixed frequency.

The frequency coherence bandwidth can be measured by cross correlating the channel transfer functions between two different frequencies denoted by its frequency index $k1$ and $k2$,

$$\rho[(k2-k1)\Delta f] = \left\langle \frac{\left| \sum_m H^*(t_m,f_{k1})H(t_m,f_{k2}) \right|}{\sqrt{\sum_m |H(t_m,f_{k1})|^2} \sqrt{\sum_m |H(t_m,f_{k2})|^2}} \right\rangle, \quad (6)$$

where the correlation is done for each packet (summing over j for a fixed n), and then averaged over all the packets (of a duration approximately 1 h). One finds that $\rho<0.2$ when $k1\neq k2$; indicating that the channel transfer functions are uncorrelated between (adjacent) frequency bins. In other words, the frequency coherence bandwidth is <80 Hz.

The probability distribution of the envelope amplitude $|\tilde{H}(t_m,f_k)|$ is determined for each frequency f_k. The results are plotted in Fig. 3 for different frequencies. One finds that

the probability distributions are very similar—suggesting that the envelope amplitudes at different frequencies (within the band) have independent and identical distributions (iid). [In Fig. 3, three frequencies bins have a slightly different distribution than the rest of the frequencies bins. This difference could be easily caused by a small number of events in the high-tail distribution (due to instantaneous interference between the signal and noise) that would shift the probability distribution to what is shown.]

Assuming the iid property, the envelope amplitudes of all frequencies (within the band) are used to increase the statistical samples. The resulting statistical distribution is plotted in Fig. 4(a) (for the 80 Hz frequency bin data).

III. THEORETICAL MODELS AND COMPARISON WITH DATA

In this section, previous well-known amplitude distribution statistics are reviewed and compared with the measured data. For this comparison, one first considers the envelope statistics at a fixed frequency.

A narrowband signal can be represent by $p(f) = H(f)e^{i2\pi f}$, where $H(f)$ is the complex amplitude

$$H(f) = X + iY, \quad (7)$$

where X and Y are often referred to as the in-phase and quadrature components of the signal. Assuming that X and Y are Gaussian random variables with a probability distribution given by

$$p(A) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(A-\langle A\rangle)^T\Sigma^{-1}(A-\langle A\rangle) \right], \quad (8)$$

where $A\equiv[X,Y]^T$, with the superscript T denoting transpose, $\langle A\rangle$ denotes the mean

$$\langle A\rangle \equiv \left\langle \begin{bmatrix} X \\ Y \end{bmatrix} \right\rangle = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad (9)$$

and Σ is the covariance matrix given by

$$\Sigma = \langle(A-\langle A\rangle)(A-\langle A\rangle)^T\rangle = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}. \quad (10)$$

Defining

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}, \quad (11)$$

it can be shown, after some calculations, that the (envelope) amplitude $Z=\sqrt{X^2+Y}$ has a distribution given by

$$p(z) = A(z,\mu_x,\mu_y,\sigma_x,\sigma_y,\rho_{xy}) \cdot \Psi(z,\mu_x,\mu_y,\sigma_x,\sigma_y,\rho_{xy}), \quad (12)$$

where
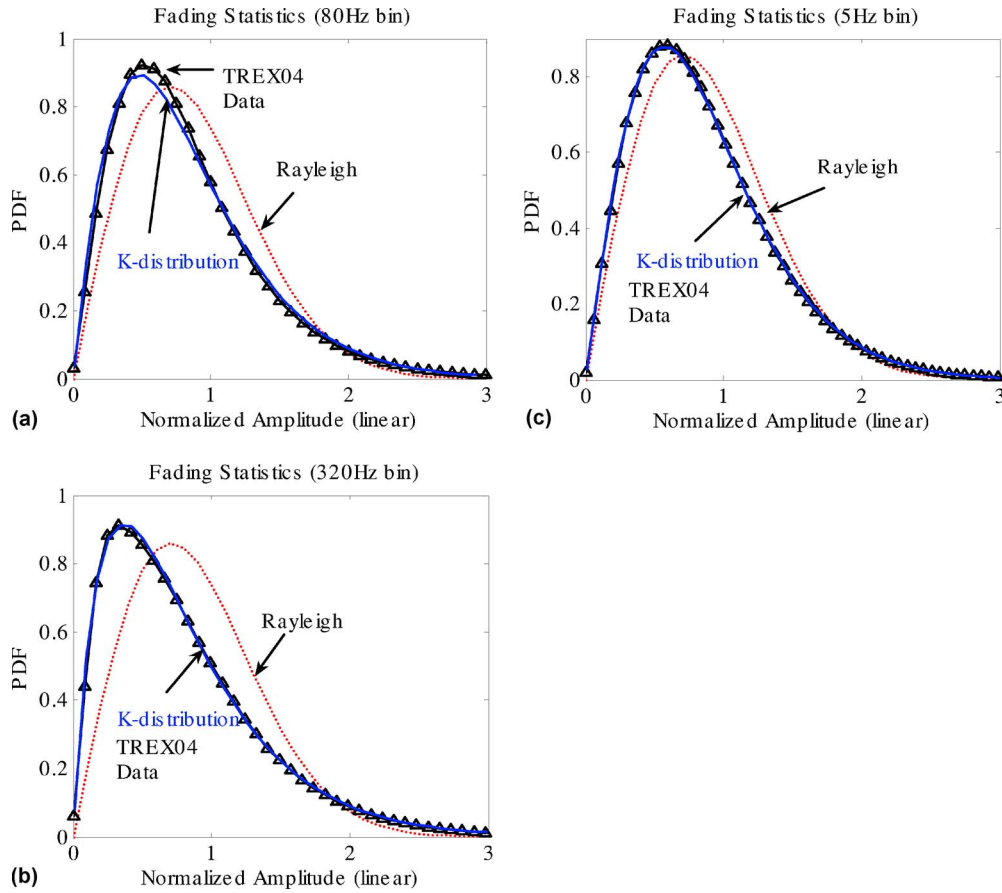
FIG. 4. (Color online) Data of Fig. 3 averaged over frequencies, shown with triangles compared with various theoretical distributions: K-distribution, Rayleigh distribution, and Mikhalevsky's model. The distribution using Mikhalevsky's model turns out to be nearly identical to Rayleigh distribution and is not separately plotted. (a) Frequency bin size=80 Hz; (b) frequency bin size=320 Hz; and (c) frequency bin size=5 Hz.

$$A(z,\mu_x,\mu_y,\sigma_x,\sigma_y,\rho_{xy}) = \frac{z}{\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}\cdot\exp\left\{\frac{-1}{2(1-\rho_{xy}^2)}\right.$$
$$\times\left[\frac{z^2+2\mu_x^2}{2\sigma_x^2}-\frac{2\rho_{xy}\mu_x\mu_y}{\sigma_x\sigma_y}\right.$$
$$\left.\left.+\frac{z^2+2\mu_y^2}{2\sigma_y^2}\right]\right\}, \quad (13)$$

$$\Psi(z,\mu_x,\mu_y,\sigma_x,\sigma_y,\rho_{xy}) = \frac{1}{2\pi}\int_0^{2\pi}\exp\left\{\frac{-1}{2(1-\rho_{xy}^2)}\right.$$
$$\times[a\cos 2\phi + b\sin 2\phi$$
$$\left.+ c\cos\phi + d\sin\phi]\right\}d\phi, \quad (14)$$

where a, b, c, and d are defined by

$$a = z^2\left(\frac{1}{2\sigma_x^2}-\frac{1}{2\sigma_y^2}\right), \quad b = z^2\left(\frac{\rho_{xy}}{\sigma_x\sigma_y}\right),$$

$$c = z\left[\frac{2\rho_{xy}\mu_y}{\sigma_x\sigma_y}-\frac{2\mu_x}{\sigma_x^2}\right], \quad d = z\left[\frac{2\rho_{xy}\mu_x}{\sigma_x\sigma_y}-\frac{2\mu_y}{\sigma_y^2}\right]. \quad (15)$$

To interpret the above amplitude distributions in the context of signal propagation, the simplest example will be direct path propagation between a source and a receiver. For this case, μ is the (deterministic) signal and $\sigma_x^2=\sigma_y^2=\sigma_N^2$ is the variance of the noise. For propagation through a random medium, μ is the deterministic component of the signal and $\sigma_x^2=\sigma_y^2=\sigma_s^2+\sigma_N^2$ is the sum of the variance of (the random component of) the signal and that of the noise.

Typical underwater acoustic propagation involves many multipaths. In that case, the complex amplitudes are summation of many multipaths, $X=\Sigma_{i=1}^N x_i$, $Y=\Sigma_{i=1}^N y_i$. This case has been studied by Mikhalevsky.[12] He hypothesized that the signal fluctuation is manifested mainly in the signal phase, as defined by $x_i+iy_i=|a_i|e^{i\phi_i}$. Assuming that the variance in phase fluctuation is the same for all multipaths, $\sigma_{\phi_i}^2=\sigma_\phi^2$, and hence $\sigma_{x_i}^2=\sigma_x^2$, $\sigma_{y_i}^2=\sigma_y^2$, the signal envelope amplitude statistics can be grouped into three cases: For unsaturated propagation, $\sigma_\phi=0$, $\mu_x\neq 0$, $\mu_y\neq 0$, $\sigma_x^2=\sigma_y^2=\sigma_N^2$, one finds $\rho_{x,y}=0$, and the envelope amplitude statistics distribution is Rician. For saturated propagations, $\sigma_\phi\to\infty$, one finds $\mu_x\to 0$, $\mu_y\to 0$, $\sigma_x^2=\sigma_y^2=\sigma_N^2$, and $\rho_{x,y}\to 0$. Thus, the envelope amplitude fluctuation approaches the Rayleigh distribution. For partially saturated propagation, which corresponds to $\sigma_\phi\neq 0$, $\mu_x\neq 0$, and $\mu_y\neq 0$, the envelope amplitude distribution depends on σ_ϕ and other parameters of the signal.

It should be pointed out that the above expressions, Eqs. (12)–(15), for the (envelope) amplitude distribution are very general results; which are independent of the propagation model. The only assumptions are that the in-phase and

TABLE I. Mean and variance measured from data.

Frequency bin (Hz)	$\mu_x(\times 10^{-4})$	$\mu_y(\times 10^{-4})$	σ_x^2	σ_y^2	$\rho_{xy}(\times 10^{-4})$
320	−0.501	−0.706	0.5007	0.4993	−0.795
80	−0.925	−1.473	0.5001	0.4999	−0.668
5	−1.636	3.209	0.4998	0.5002	4.857

quadrature components (X and Y) are Gaussian random variables. For data analysis, one can measure the mean and variance of X and Y and their correlation directly from the data $H(f_k)$. Using the TREX04 data analyzed above, the measurement results are given in Table I for three different signal designs (different symbol durations or different frequency bin sizes). Note that in Table I $\rho_{xy} \ll \sigma_x$, σ_y and $\rho_{xy} \approx 0$. As remarked above, for $\rho_{xy} \to 0$, Eqs. (12)–(15), reduce to the Rayleigh or Rician distribution depending on whether or not $\mu^2 = \mu_x^2 + \mu_y^2$ is zero. For the data analyzed here, one notes that $\mu^2 \ll \sigma^2$ (see Table I), and can thus assume $\mu^2 = 0$. It then yields the Rayleigh distribution. In general, $\mu^2 \neq 0$ and one then has the Rician distribution. (The Rayleigh and Rician distributions are widely used in the literature for modeling the performance of communications in the RF world.)

The Rician (amplitude) distribution is given by[15]

$$p_z(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2 + \mu^2}{2\sigma^2}\right) I_o\left(\frac{\mu z}{\sigma^2}\right), \quad z \geq 0, \qquad (16)$$

where z is the amplitude and I_o is the modified Bessel function of the first kind of zero order. Equation (16) is obtained from Eq. (12) when $\rho_{x,y} = 0$, $\mu^2 = \mu_x^2 + \mu_y^2$, and $\sigma_x^2 = \sigma_y^2 = \sigma^2$. The first and second moments of z have been calculated[15]

$$E\{Z\} = \sqrt{\frac{\pi}{2}} \sigma \exp(-\bar{K}/2)[(1 + \bar{K}) I_o(\bar{K}/2) + \bar{K} I_1(\bar{K}/2)], \qquad (17)$$

$$E\{Z^2\} = 2\sigma^2 + \mu^2 = 2\sigma^2(1 + \bar{K}), \qquad (18)$$

where $\bar{K} = \mu^2/2\sigma^2$ is referred to as the Rician parameter (known as the power ratio of specular-to-scatter or deterministic-to-random signal). I_1 is the modified Bessel function of the first kind of first order. A special case of the Rician distribution, for $\mu = 0$ (or $\bar{K} = 0$), is the well-known Rayleigh distribution,

$$p_z(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad z \geq 0. \qquad (19)$$

The first and second moments of the Rayleigh distribution are given by

$$E(Z) = \sqrt{\pi/2} \cdot \sigma \qquad (20)$$

and

$$E(Z^2) = 2\sigma^2. \qquad (21)$$

Given the signal parameters measured from the data (Table I, 80 Hz data), it is simple to check if the Rayleigh or Rician model fits the high-frequency underwater acoustic data (the envelope amplitude statistics) reported above [Fig.

4(a)]. The Rayleigh distribution using the parameter values, given in Table II, is shown in Fig. 4(a) to compare with the data. Also shown is the amplitude distribution using Eq. (12), which is practically the same as the Rayleigh distribution. This is expected, since $\bar{K} = \mu^2/2\sigma^2 \sim O(10^{-8})$ (see Table I).

Equation (12), derived with a very general assumption that the in-phase and quadrature components are Gaussian random variables, does not seem to fit the data. (Of course, neither does the Rayleigh distribution.) This conclusion is also checked using a different signaling design, by varying the frequency bin size Δf from 80 Hz to 320 Hz, and to 5 Hz. The measured envelop amplitude distributions and the modeled probability distributions are plotted in Figs. 4(b) and 4(c). These plots again show the disagreement between the model and the data. The mean and variance of the data are summarized in Tables I and II.

It should be noted that the Rician distribution will yield a worse fit to the data since the Rician distribution has a higher percentage of high amplitude data points than the Rayleigh distribution, whereas the data indicate a higher percentage of low amplitude data points than the Rayleigh distribution [Fig. 4(a)] One concludes that the amplitude statistics for high-frequency underwater communication signals is neither Rician nor Rayleigh distribution, as in RF channels.

Next, the discussion will be focused on a popular non-Rayleigh model called the K-distribution, which seems to fit the data very well. In this section, the K-distribution model will be presented and fitted with the data. The motivation, derivation, and interpretation of this model will be given in the next section.

The K-distribution is given by[16]

$$p_z(z) = \frac{4}{\sqrt{\alpha}\Gamma(v)}\left(\frac{z}{\sqrt{\alpha}}\right)^v K_{v-1}\left(\frac{2z}{\sqrt{\alpha}}\right), \quad z \geq 0, \qquad (22)$$

where v is a shape parameter, α is a scale parameter, K_{v-1} is the modified Bessel function of the second kind, of order $v - 1$, and $\Gamma(v)$ is the Gamma function. The first and second moments of the K-distribution are as below:

TABLE II. Parameters for Rayleigh and K-distribution.

Frequency bin (Hz)	K-distribution		Rayleigh
	v	α	σ_r^2
320	1.2850	0.7787	0.5003
80	2.1290	0.4700	0.5003
5	3.7110	0.2695	0.5001

$$E(Z) = \frac{\sqrt{\alpha\pi}}{2} \cdot \frac{\Gamma(v+0.5)}{\Gamma(v)} \qquad (23)$$

and

$$E(Z^2) = \alpha v. \qquad (24)$$

A special case of the K-distribution, as $v \rightarrow \infty$ and $\alpha v = 2\sigma^2$ remains constant, is a Rayleigh distribution given in Eq. (19).

The model prediction using the K-distribution is shown in Figs. 4(a)–4(c) to compare with the data. The parameters for the K-distribution are determined from the mean and variance of the data using Eqs. (23) and (24). Their measured values are given in Table II. One finds fairly good agreement between the model prediction and data.

IV. LONG-TERM AND SHORT-TERM FLUCTUATION STATISTICS

To guide the motivation of the analysis in this section, and the interpretation of the derived model, it helps to point out the main difference between the current work and the previous work. The main difference lies in the underlying mechanism for signal fluctuations (between low and high frequencies) that lead to the Rayleigh/Rician model, on the one hand; and the K-distribution model, on the other hand.

One recalls that an implicit assumption used in deriving Eq. (12) is that the random variables X and Y follow stationary Gaussian statistics. This assumption seems to be valid for low-frequency signal propagation, but perhaps not appropriate for high-frequency signal propagation. High-frequency signals may follow quasi-stationary statistics that involve two time scales associated with long-term fading and short-term fading.[2] Over a short time scale, the high-frequency signal is heavily influenced by the microfine structures (e.g., turbulence) in the ocean. The signal amplitude fluctuation follows a short-term statistics. Over a long time scale, the amplitude fluctuations of the signal will likely be dominated by the fine-structure perturbations of the ocean, assuming that the rapid fluctuations induced by the microstructures have been averaged out. The signal amplitude fluctuations follow long-term statistics, which may be different from the short-term statistics. At the symbol level, the symbol amplitude envelope statistics follow a joint probability distribution, determined by the short-term probability distribution function conditioned on the amplitude distributions dictated by the long-term probability distribution function. This approach will be used in this section. We discuss the theoretical framework, determine the long- and short-term statistics from the data, and show that the overall statistics follow a distribution—numerically very close to the K-distribution.

[Note that the microfine structures have a negligible influence on low-frequency signals and thus can be ignored. Consequently, there is only one time scale for low-frequency signal fluctuation.]

To obtain the long-term statistics, the data will be averaged over a time period of T, such that the short-term signal fluctuation has been averaged out. In the time domain, the received signal is given by

$$r(t) = h(t) \otimes s(t). \qquad (25)$$

Let time be segmented into multiple periods of T,

$$t = nT + j\Delta t \equiv T_n + \tau_j, \qquad (26)$$

where j goes from 1 to $T/\Delta t$. Assuming that $\bar{h}(T_n)$ is the mean (average) channel impulse response over the time period between T_n and T_{n+1}, the long-term statistics are determined by $\bar{h}(T_n)$ for different n. For a given time window (between T_n and T_{n+1}), one can measure the short-term signal fluctuations (during this time window) using

$$\tilde{r}(t) = \bar{h}^{-1}(T_n) \otimes h(t) \otimes s(t), \qquad (27)$$

where $\bar{h}^{-1}(T_n)$ is an inverse filter (as a function of delay time) estimated at geotime T_n.

[The meaning of Eq. (27) can be illustrated with other familiar examples. One notes that Eq. (27) is similar to passive time-reversal (phase conjugation) operation, except that the inverse of the channel response is used instead of the conjugated time-reversed impulse response. Another usage of Eq. (27) will be for a channel in which the signal consists of deterministic eigenray arrivals plus many time-varying micropaths. Equation (27) allows one to remove the mean channel effect (the deterministic eigenrays) so that the fluctuation statistics of the micropaths can be studied.]

The residual impulse response that characterizes the short-term fluctuation is given by

$$\tilde{h}(\tau, T_n) = \bar{h}^{-1}(T_n) \otimes h(t). \qquad (28)$$

The above equation is expressed in the frequency domain as

$$\tilde{H}(f_k, \tau)\big|_{T_n} = H(f_k, t)/\bar{H}(f_k, T_n). \qquad (29)$$

A. Long-term envelope amplitude statistics

The long-term fluctuation statistics is obtained by averaging the signal intensity $|H(f_k, t)|^2$ over a period of T, and taking the square root to obtain one realization of $|\bar{H}(f_k, T_n)|$. The data are processed using $T = 0.2$ and 0.4 s; the results are practically the same. Henceforth, $T = 0.2$ will be used unless otherwise mentioned. How to determine the appropriate time scale T will be addressed later in Sec. IV D. For $T = 0.2$ s, one has 7169 realizations of the signal intensities for each frequency (out of 134 packets).

One finds that $|\bar{H}(f_k, T_n)|$ approximately follows a lognormal distribution. It is thus appropriate to plot the distribution in terms of $y = 20 \log |\bar{H}(f_k, T_n)|$ for a fixed f_k. Figure 5(a) plots the histogram distribution for $20 \log |\bar{H}(f_k, T_n)|$ for different f_k. Note that the distributions for different frequencies (within the band) are approximately the same; they are independently distributed as remarked above. Figure 5(b) also shows the distribution averaged over f_k within the signal bandwidth of 4 kHz.

To model the data, the lognormal distribution given by

2622   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

W. Yang and T. C. Yang: High frequency channel characterization

(a)

(b)
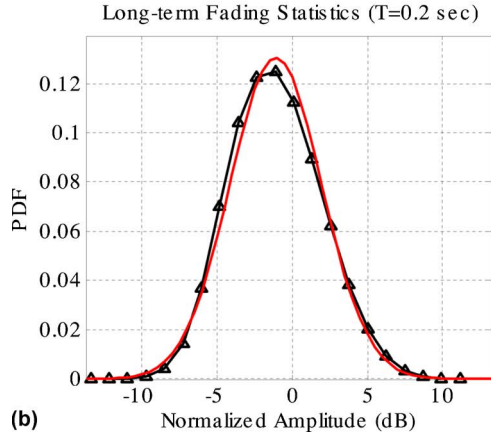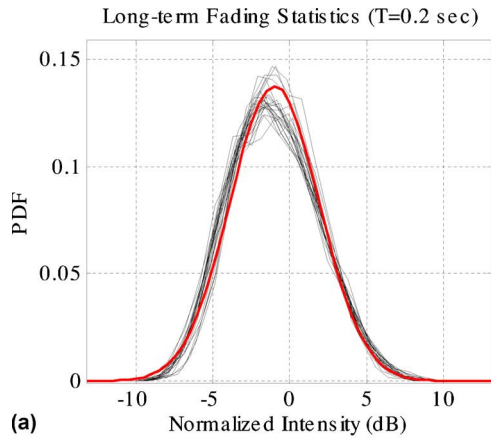
FIG. 5. (Color online) Long-term fading statistics ($T=0.2$ s) vs a Gaussian distribution in (dB). (a) Thin lines are distributions for individual frequencies. (b) The line with triangles is for the average over all frequency bins. The smooth thick line in both figures is a fit to the data using the lognormal distribution.

(a)

(b)

FIG. 6. (Color online) Short-term fading statistics ($T=0.2$ s) vs the Rayleigh distribution. (a) Thin lines are distributions for individual frequencies. (b) The line with triangles is for the average over all frequency bins. The smooth thick line in both figures is a fit to the data using the Rayleigh distribution.

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} x^{-1} \exp\left(-\frac{(\ln x - \mu_x)^2}{2\sigma_x^2}\right), \tag{30}$$

is used, where x is $|\bar{H}(f_k, T_n)|$, for $n=1,\ldots N$. Let $u = 20\log_{10}(x)$, the probability distribution for u is

$$p(u) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left(-\frac{(u - u_u)^2}{2\sigma_u^2}\right), \tag{31}$$

which is a Gaussian distribution with a mean and variance given by

$$\mu_u = 20\mu_x/\ln 10, \quad \text{and} \quad \sigma_u = 20\sigma_x/\ln 10. \tag{32}$$

One finds that the data in Fig. 5(b) can be fitted with Eq. (31) with $\mu_u = -1.0968$ dB, and $\sigma_u^2 = 9.3468$ dB^2. The distribution of $|\bar{H}(f_k, T_n)|$ has a mean, $\mu_x = -0.128$, and a variance of $\sigma_x^2 = 0.128$ for the digitized data. (Unit of digitization is arbitrary.)

B. Short-term envelope amplitude statistics

The short-term distribution is obtained from data using Eq. (29) for each period of T and then summed/averaged over all periods of data. Again, $T=0.2$ and 0.4 s are used, and the results remain practically the same. Using $T=0.2$ s, the resulting distribution is plotted in Fig. 6(a) for individual

frequency bins. There are 16 realizations for each 0.2 s segment; hence, a total of $16 \times 7169 = 114\,704$ realizations for the short-term statistics for each frequency component. One observes that the majority of frequencies have near-identical distributions. A few frequency components have slightly shifted distributions. They are the manifestation of the shifted distributions in Fig. 3. Invoking the assumption that the distributions for various frequencies are independent and identically distributed, the distributions are averaged over frequencies to obtain a more statistically meaningful measurement. The result is plotted in Fig. 6(b). The data in Fig. 6(b) are well fitted with a Rayleigh distribution [Eq. (19)] in terms of y. It has a mean of $\mu_y = 0.8874$ and a variance of $\sigma_y^2 = 0.5013$.

C. Modeling the overall statistics

Given a long-term distribution of lognormal and a short-term distribution of Rayleigh, the resulting distribution can be calculated and will be shown to be almost like a K-distribution.

For mathematical convenience, consider a fixed frequency, and use the following notation: $x \equiv \bar{H}(f_k, T_n)$, $y \equiv \tilde{H}(f_k, \tau_j)$ and $z \equiv H(f_k, t)$. Equation (29) then becomes

$$z = xy. \tag{33}$$

The probability distribution for z, as x and y are mutually independent, is given by

$$p_Z(z) = \int p_{Z|X}(z|X=x)p_X(x)dx = \int p_{z,x}(z,x)dx$$

$$= \int p_{Y,X}(y=z/x,x)\frac{1}{x}dx = \int p_Y(z/x)p_X(x)\frac{1}{x}dx \tag{34}$$

where the first line of Eq. (34) states that the overall distribution of z is equal to the short-term probability distribution, conditioned on the long-term probability distribution and integrated over the long-term probability distribution. Assuming that the short-term and long-term distributions are independent, it becomes the joint probability distribution in z, x, and is transformed by change of variables to y, x in the second line of Eq. (34).

Substituting the lognormal distribution, Eq. (30), for $P_X(x)$ and the Rayleigh distribution, Eq. (19), for $P_Y(y)$, one finds, after some algebra,

$$p_Z(z) = \int \left(\frac{1}{\sqrt{2\pi}\lambda}\frac{1}{\nu}e^{(\ln \nu - \mu)^2/2\lambda^2} \right) \left(\frac{z}{\nu}e^{-z^2/2\nu} \right)d\nu, \tag{35}$$

where

$$\nu = \sigma_y^2 x^2, \quad \lambda = 2\sigma_x, \quad \mu = 2\ln \sigma_y + 2\mu_x, \tag{36}$$

The K-distribution is a mixture of Rayleigh and Gamma distributions and is defined as below:

$$p_Z(z) = \int \left(\frac{\nu^\beta}{(2\alpha)^{1+\beta}\Gamma(1+\beta)}e^{-\nu/\alpha} \right) \left(\frac{z}{\nu}e^{-z^2/2\nu} \right)d\nu$$

$$= \frac{4}{\sqrt{\alpha}\Gamma(1+\beta)} \left(\frac{z}{\sqrt{\alpha}} \right)^{1+\beta} K_\beta\left(\frac{2z}{\sqrt{\alpha}} \right),$$

$$z \geq 0, \quad \alpha > 0, \quad \beta > -1, \tag{37}$$

where α and β are parameters of the Gamma distribution. It has been proven that lognormal and Gamma distributions are close approximations of each other.[17,18] Consequently, one finds that Eqs. (35) and (37) are numerically close approximations of each other with the following relation between the parameters[19]

$$\mu = \ln(\alpha/2) + \Psi(1+\beta), \quad \lambda^2 = \Psi'(1+\beta),$$

where Ψ and Ψ' are the psi-function and its derivative.

For a consistency check, note that the measured parameters from the long-term statistics are: $\mu_x = -0.128$ and $\sigma_x^2 = 0.128$, and the parameters for short-term statistics are: $\mu_y = 0.8874$, $\sigma_y^2 = 0.5013$. This yields $\mu = -0.9466$ and $\lambda = 0.7155$ based on Eq. (36). From data fitting in Sec. III C (Table II, 80 Hz data), one finds $\alpha = 0.470$, and $\beta = \nu - 1 = 1.129$, which yields $\mu = -0.9454$ and $\lambda = 0.7724$. Thus, both methods approximately yield the same answer.

The above analysis showed that the overall data distribution, based on (the joint probability distributions of) the long-term (lognormal) and short-term (Rayleigh) distributions, follow a distribution which is numerically close to the K-distribution. In numerical modeling of data (Sec. III), the

K-distribution is used because it is a well-established special function, whereas there is not a named special function for Eq. (35). Since all models are numerical fit of the data, for data analysis, it is more convenient to use K-distribution as a substitute.

D. Determination of the time scale *T*

The previous subsections use $T=0.2$ s. Naturally, the question is what is the time scale separating the short term from the long term fluctuations. The time scale T is, in principle, related to the time scale of the appropriate oceanographic processes, which dominate the signal fluctuations at the corresponding frequency domain. The microfine structures of the ocean inhomogeneities (ocean turbulence) have a time scale of milliseconds to seconds, and mostly affect only the high-frequency acoustic signal propagations. The fine structures of the ocean inhomogeneities cover a time scale from seconds to hours. The high-frequency end (corresponding to a time scale of seconds) of the fine structure spectrum affects the high-frequency acoustic signal propagations more, as the wavelengths of the inhomogeneities match more closely with the acoustic wavelengths. The low-frequency part of the inhomogeneities affects the low-frequency acoustic propagations more.

The acoustic signal propagation theory has suggested that the envelope amplitude statistics for fully saturated fluctuations should be Rayleigh distributed. This is expected for a random medium of microfine structures. Indeed, the short-term statistics is well fitted by the Rayleigh distribution. This suggests that the time scale T is related to the time scale of microfine structures. The long-term statistics are then associated with the fine-structure processes. The observed lognormal distribution is consistent with the fact that the distribution is partially saturated.

While there is a qualitative understanding of the effect of the ocean inhomogeneities on acoustic signal propagations, there is no theoretical derivation of T based on the random media propagation theory. In this section, an empirical estimation of T will be pursued based on the "best" fit of the statistical distributions to the data. Note that changing the value of T will affect both the long-term and short-term distributions. Data analysis results indicate that if one uses T significantly different from 0.2 s, the parameters of the overall distribution—deduced from the long-term and short-term statistics—do not agree so well with the parameters of the K-distribution deduced directly from data. This indicates that there is a preferred value for the time scale T that separates the short-term fluctuations from the long-term fluctuations.

The method used to determine T is based on the "goodness of fit" measure of the root-mean-squared error (RMSE), also known as a fit standard error. The RMSE is defined as

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{L}\sum_{i=1}^{n} (F_i - S_i)^2} \tag{38}$$

where $L = n - m$ indicates the number of independent pieces of information involving the n data points and m parameters of the prospective probability distribution. S_i denotes the sampled statistical function (either the probability distribu-

W. Yang and T. C. Yang: High frequency channel characterization
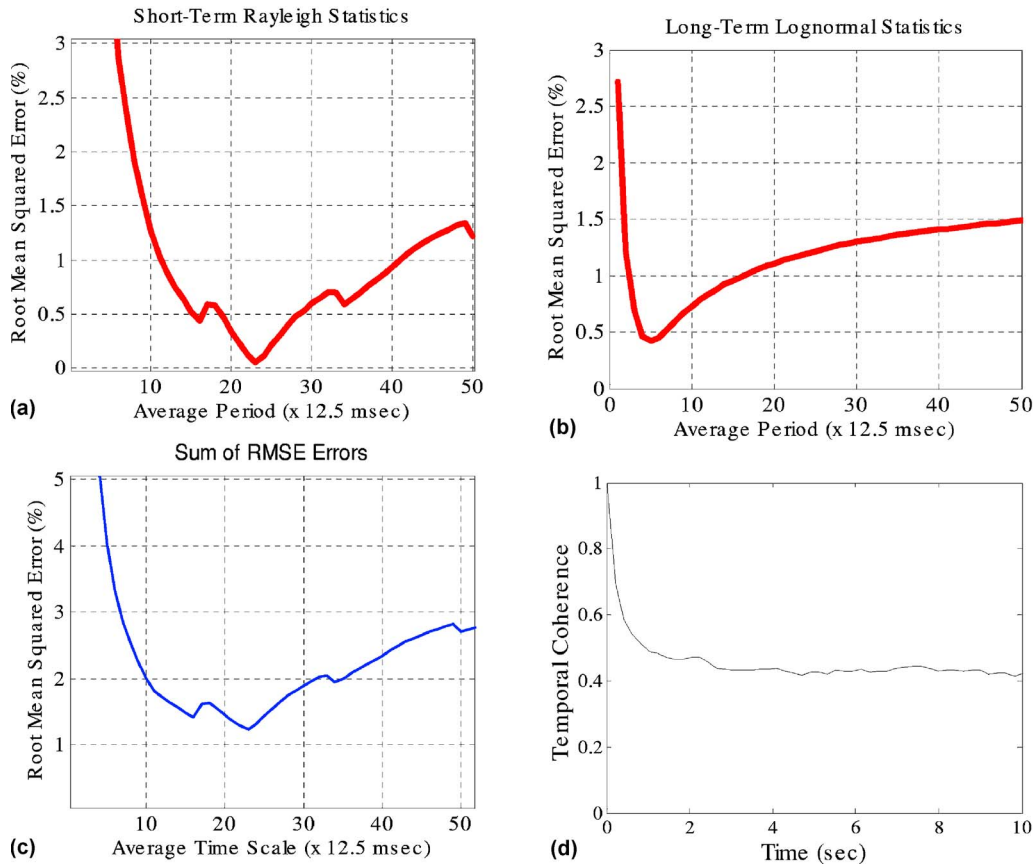
FIG. 7. (a) RMSE between short-term statistics of experimental data and the Rayleigh distribution. (b) RMSE between long-term statistics of experimental data and the lognormal distribution. (c) Sum of long-term and short-term RMSE. (d) The signal temporal coherence vs lag time averaged over >50 packets of data.

tion or the cumulative probability distribution function) based on data at amplitude x_i, $i=1,\ldots,n$. F_i denotes the sampled statistics at amplitude x_i based on the statistical model. For the above data, $L \sim 3 \times 10^5$ for the long-term statistics and $L \sim 5 \times 10^6$ for the short-term statistics.

Equation (38) will be evaluated for different values of T. For each value of T, one obtains the short-term and long-term fluctuation statistics from the data. The short-term and long-term data are fitted with the Rayleigh and lognormal distributions, respectively. The parameters of the distributions are obtained by the best fit with the data. Having determined the parameters for the best fit, one determines the RMSE of the fit using Eq. (38).

Consider the short-term envelope amplitude statistics first. Figure 7(a) shows the RMSE values of Rayleigh statistics as a function of the time scale T. The minimum of the test is located at $T \approx 0.3$ s. One finds that $T = 0.2$ and 0.4 s yield a RMSE$<1\%$. In fact, the RMSE is $\leqslant 1\%$ for T between 0.15 and 0.5 s. It indicates that the Rayleigh distribution is a good fit for the short-term fading statistics during this time window. When T exceeds 0.5 s, the fit deteriorates significantly. This result suggests that for $T>0.5$ s the effects of fine-structure processes are no longer negligible and need to be included.

For long-term statistical distribution, the RMSE between the data and the lognormal distribution is plotted in Fig. 7(b). One finds that the RMSE is $\leqslant 1\%$ for T between 0.04 and 0.21 s, and $\leqslant 1.5\%$ for T between 0.03 and 0.63 s. In fact,

the RMSE remains below 2% for a large window of T [to at least 1.25 s in Fig. 7(b)]. This is consistent with the expectation that the long-term fluctuation should not change significantly with the time window T as long as it is "long term." One notes that there is no theoretical basis that the distribution has to be lognormal, which may just be a convenient choice. Thus, a 2% RMSE is quite reasonable. Figure 7(c) plots the sum of the short-term and long-term RMSE. The minimum occurs around $T \sim 0.25$ s. One finds that $T = 0.2$ and 0.4 s yield a reasonable RMSE.

Figure 7(d) shows the temporal coherence of M-sequence signals determined by the correlation of the first M-sequence signal with the following M-sequence signals.[2] One temporal coherence curve is obtained for one packet. The coherence curves are averaged over many (50) packets. The result is shown in Fig. 7(d). It indicates that the signal coherence time is of the order of 0.2–0.4 s, after that the coherence value goes below 0.5. This result suggests that the signal coherence time can be used as the time scale separating the short-term from the long-term fluctuations, as in this case (high-frequency propagation) both are determined by the microfine (turbulence) disturbances. Whether this is a good assumption remains to be investigated.

We have also evaluated the fit of the data with the theoretical distribution using the Kolmogorov-Simirnov test statistics.[16] The results are very similar and not explicitly shown here.

V. CONCLUSIONS

In this paper, narrowband envelope amplitude (fading) statistics are investigated for high-frequency signals over a wide band of frequencies (15–19 kHz). The envelope amplitude statistics shows a non-Rayleigh or a non-Rician distribution behavior. The conventional models for the envelope amplitude distributions, developed for low-frequency unsaturated, partially saturated, and fully saturated signal fluctuations; do not fit the high-frequency amplitude statistics data. The reason is that these models assume a fading statistics that is valid for all time scales. The above analysis of the high-frequency data indicates two time-scale fading phenomena: Long-term versus short-term fading. The division between the two is determined by using RMSE test, which is about a 0.2 s time scale for the TREX04 data.

It is found that the long-term amplitude fading statistics follow a lognormal distribution, and the short-term amplitude fading statistics follows a Rayleigh distribution. The signal amplitude distribution based on the joint long-term and short-term distributions yields a distribution numerically close to the K-distribution, which is found to be a good fit of the high-frequency data.

ACKNOWLEDGMENTS

[1] J. G. Proakis, *Digital Communications* (McGraw–Hill, New York, 1995).

[2] T. C. Yang, "Measurements of temporal coherence of sound transmissions through shallow water," J. Acoust. Soc. Am. **120**(5), November 2006.

[3] R. Galvin and L. S. Wang, "Measured channel characteristics and the corresponding performance of an underwater acoustic communication system using parametric transduction," *IEE Proceedings—Radar, Sonar and Navigation*, Vol. 147, 247–253 (2000).

[4] X. Geng and A. Zielinski, "An eigenpath underwater acoustic communication channel model," *Proceedings of IEEE Oceans '95 Conference*, 1189–1196 (1995).

[5] L. E. Estes, G. Fain, and D. Carvalho, "Simulation of a shallow water acoustic telemetry channel," *Proc. IEEE Oceans '92 Conference*, 584–589 (1992).

[6] W. Jobst and X. Zabalgogeazcoa, "Coherence estimates for signals propagation through acoustic channels with multiple paths," J. Acoust. Soc. Am. **65**, 622–630 (1979).

[7] R. Galvin and R. F. W. Coates, "A stochastic underwater acoustic channel model," *Proceedings of IEEE Oceans '96 Conference*, 203–210 (1996).

[8] R. J. C. Bultitude, "Measurement, characterization, and modeling of indoor 800/900 MHz radio channels for digital communications," IEEE Commun. Mag. 5–12 (June 1987).

[9] W. C. Y. Lee, *Mobile Communications Engineering* (McGraw–Hill, New York, 1993).

[10] T. S. Rappaport, "Indoor radio communications for factories of the future," IEEE Commun. Mag. 15–24 (May 1989).

[11] S. M. Flatte *et al.*, *Sound Transmission Through a Fluctuating Ocean*, edited by S. M. Flatte, (Cambridge University Press, Cambridge, UK, 1979).

[12] P. N. Mikhalevsky, "Envelope statistics of partially saturated processes," J. Acoust. Soc. Am. **72**, 151–158 (1982).

[13] J. A. Capivovic, A. B. Baggeroer, K. Von Der Heydt, and D. Koelsch, "Design and performance analysis of a digital telemetry system for short range underwater channel," IEEE J. Ocean. Eng. **9**, 242–253 (1984).

[14] W.-B. Yang and T. C. Yang, "M-ary frequency-shift-keying communications over an underwater acoustic communication channel: Performance comparison of data with models" (Nov. 2006).

[15] S. O. Rice, "Mathematical analysis of random noise," Bell Syst. Tech. J. **23**, 98–101 (1945).

[16] D. A. Abraham, "Modeling non-Rayleigh reverberation," *SR-266*, SACLANT Undersea Research Center, La Spezia, Italy, 1997.

[17] J. R. Clark and S. Karp, "Approximations for lognormally fading optical signals," Proc. IEEE **58**, 1964–1965 (1970).

[18] N. L. Johnson and S. Kotx, *Distributions in Statistics: Continuous Univariate Distributions* (Wiley, New York, 1970).

[19] A. Abdi and M. Kaveh, "K-distribution: An appropriate substitute for Rayleigh-lognormal distribution in fading-shadowing wireless chanels," Electron. Lett. **34**, 851–852 (1998).

Multiple source localization using a maximum *a posteriori* Gibbs sampling approach

Zoi-Heleni Michalopoulou[a]
Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey 07102

Multiple source localization in underwater environments is approached within a matched-field processing framework. A maximum *a posteriori* estimation method is proposed that estimates source location and spectral characteristics of multiple sources via Gibbs sampling. The method facilitates localization of weak sources which are typically masked by the presence of strong interferers. A performance evaluation study based on Monte Carlo simulations shows that the proposed maximum *a posteriori* estimation approach is superior to simple coherent matched-field interference cancellation. The proposed method is also tested on the estimation of the number of sources present, providing probability distributions in addition to point estimates for the number of sources. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354027]

I. INTRODUCTION

Acoustic source localization in shallow water environments is a challenging task hampered by the effects of sound interactions with the inadequately understood propagation medium. Loud interfering sources often present in coastal waters mask the source of interest, further complicating localization.

Matched field processing (MFP)[1,2] has been extensively used for source localization in the ocean. The measured acoustic data are compared to field replicas generated with a sound propagation model and estimates are taken to be those parameter values that generate the best match between true and replica fields according to some specified matching criterion.

Variants of MFP have been proposed for localization of multiple sources. Model-based methods in underwater environments were discussed in Refs. 3–8; these multiple source localization techniques typically rely on eigenvector decompositions, modified Bartlett functions, or combination of these two approaches. The CLEAN method, widely used in astronomy for denoising images, was originally proposed in Ref. 1 and subsequently in Ref. 9 for weak source localization in the ocean by canceling strong interferers. In this paper we propose a new processor that estimates multiple source locations with a maximum *a posteriori* (MAP) approach optimized with Gibbs sampling. The processor forms joint posterior probability distributions of locations of multiple sources, source spectra, noise variance, as well as number of sources that are present. It is compared and found superior to simple coherent interference cancellation for multiple source localization through a Monte Carlo performance evaluation. Our method not only produces point estimates of the unknown parameters but probability distributions as well, capturing uncertainty in the estimation problem. With sampling performed from conditional posterior probability distribu-

tions, the method searches the parameter space efficiently, concentrating on regions that generate significant probability.

The structure of the paper is as follows: Section II discusses the CLEAN approach as it pertains to simple matched-field interference cancellation, or, equivalently, multiple source localization using conventional MFP. Section III describes the new processor, deriving conditional posterior distributions that are necessary for the operation of the Gibbs sampler. Section IV compares the proposed processor to a simple coherent cancellation processor, relying on Bartlett ambiguity surfaces and subtraction of strong field components, through a performance evaluation with synthetic data. Estimation of the number of sources is discussed in Sec. V. The effect of environmental mismatch on the performance of the new processor and the simple coherent canceller is addressed in Sec. VI. Section VII summarizes the work and presents conclusions.

II. MATCHED-FIELD PROCESSING AND INTERFERENCE CANCELLATION

In Ref. 1 it was suggested that the CLEAN deconvolution technique,[10,11] widely used since the late 1970s for image enhancement especially in astronomy, might be a suitable tool in localization using MFP in the presence of multiple sources. Estimates of contributions from strong interferers are subtracted from the field received at a set of sensors in an effort to reveal weaker sources of interest which are otherwise masked. Proper subtraction of field components from received signals requires accurate knowledge of how signals propagate from source to receiver. Essentially, we need to subtract from the sensed data the convolution of the source wave form and the medium impulse response, deconvolving interference from the received field.

It was proposed in Ref. 1 that, assuming a perfectly known environment, an ambiguity surface $P(r, z_s)$ (Bartlett, for example) is first constructed:

[a]Electronic mail: michalop@njit.edu

$$P(r, z_s) = \frac{\mathbf{G}(r, z_s)^* \mathbf{X} \mathbf{X}^* \mathbf{G}(r, z_s)}{\|\mathbf{G}(r, z_s)\|^2}, \tag{1}$$

where $\mathbf{G}(r, z_s)$ is a Green's function vector calculated for source range r and depth z_s, and $\mathbf{X}$ is a vector containing received data. Specifically, assuming a strong and a weak source being present:

$$\mathbf{X} = \alpha_{str} \mathbf{G}(r_{str}, z_{s,str}) + \alpha_w \mathbf{G}(r_w, z_{s,w}) + \mathbf{W}, \tag{2}$$

where α_{str} and α_w are the complex amplitudes of the two source signals, r_{str} and r_w are the ranges for the strong and weak source, and $z_{s,str}$ and $z_{s,w}$ are the corresponding depths, $\mathbf{G}(r_{str}, z_{s,str})$ and $\mathbf{G}(r_w, z_{s,w})$ are Green's function vectors calculated at the L receiving phones for the strong and weak source, respectively, and $\mathbf{W}$ is an L-dimensional complex, zero-mean noise vector.

The maximum of this ambiguity surface is expected to lead to the source location estimates $\hat{r}_{str}$ and $\hat{z}_{s,str}$, of the strongest source present. Subsequently, a new surface P_{str} is produced, using now as data the replica field computed at the estimated source location of the previous step:

$$P_{str}(r, z_s) = \frac{\mathbf{G}(r, z_s)^* \hat{\mathbf{G}}_{str} \hat{\mathbf{G}}_{str}^* \mathbf{G}(r, z_s)}{\|\mathbf{G}(r, z_s)\|^2}, \tag{3}$$

where $\hat{\mathbf{G}}_{str} = \mathbf{G}(\hat{r}_{str}, \hat{z}_{s,str})$.

This new ambiguity surface is subtracted from the original surface; the residual ambiguity surface $P_{res}(r, z_s)$ is scanned again for a new source location. The subtraction is repeated until M sources are identified, where M is the number of sources believed to be present. This "cleaning" approach cancels contributions of strong sources-interferers in an incoherent manner, because the subtraction is performed at the ambiguity surface level. In essence, we calculate $P_{res}(r, z_s)$ as follows:

$$P_{res}(r, z_s) = P(r, z_s) - k P_{str}(r, z_s), \tag{4}$$

where k is a factor depending on the strength of the loud source. A more detailed CLEANing process for MFP was developed in Ref. 9.

In our work, we followed the suggestion in Ref. 1 and used as a value for k an estimate of the strength of the loud source $|\hat{\alpha}_{str}|^2$ (similar to Ref. 9). Optimally, a maximum likelihood estimate $\hat{\alpha}_{str}$ of α_{str} under the assumption that the data are embedded in Gaussian, spatially white noise can be obtained as follows:

$$\hat{\alpha}_{str} = \frac{\hat{\mathbf{G}}_{str}^* (\mathbf{X} - \alpha_w \mathbf{G}_w)}{\|\hat{\mathbf{G}}_{str}\|^2}, \tag{5}$$

where $\mathbf{G}_w = \mathbf{G}(r_w, z_{s,w})$.

Estimation using Eq. (5) presumes knowledge of α_w and $\mathbf{G}_w$ which is not available. If, however, the strength of the weak source is much smaller than that of the strong source, that is, $|\alpha_w| \ll |\alpha_{str}|$, $\mathbf{X} \approx \alpha_{str} \mathbf{G}_{str} + \mathbf{W}$ and

$$\hat{\alpha}_{str} \approx \frac{\hat{\mathbf{G}}_{str}^* \mathbf{X}}{\|\hat{\mathbf{G}}_{str}\|^2}. \tag{6}$$

Equation (4) now becomes

$$P_{res}(r, z_s) = P(r, z_s) - |\hat{\alpha}_{str}|^2 P_{str}(r, z_s). \tag{7}$$

The process of Eq. (7) removes from the original ambiguity surface $P(r, z_s)$ some of the contributions of the strong source (provided that the estimate for the strong source power is reasonable); other contributions (namely, $\mathbf{G}^* \alpha_w \mathbf{G}_w \alpha_{str}^* \mathbf{G}_{str}^* \mathbf{G} / \|\mathbf{G}\|^2$ and $\mathbf{G}^* \alpha_{str} \mathbf{G}_{str} \alpha_w^* \mathbf{G}_w^* \mathbf{G} / \|\mathbf{G}\|^2$) are not subtracted. In an attempt to remove as comprehensively as possible the strong source contributions, we propose a coherent application of the previously described CLEAN approach that removes the strong source components at the field rather than the ambiguity surface level. A simple coherent CLEAN process is implemented as follows. Estimates of the source location $\hat{r}_{str}$ and $\hat{z}_{s,str}$ and replica $\mathbf{G}(\hat{r}_{str}, \hat{z}_{s,str})$ are obtained in the way already described. Estimate $\hat{\alpha}_{str}$ is then calculated from Eq. (6). A new ambiguity surface, $P_{res,coh}(r_w, z_{s,w})$, is calculated as follows:

$$P_{res,coh}(r, z_s) = \frac{\mathbf{G}(r, z_s)^* \mathbf{Y} \mathbf{Y}^* \mathbf{G}(r, z_s)}{\|\mathbf{G}(r, z_s)\|^2}, \tag{8}$$

where $\mathbf{Y} = \mathbf{X} - \hat{\alpha}_{str} \mathbf{G}(\hat{r}_{str}, \hat{z}_{str})$. Assuming that two sources, a strong and a weaker one, are present, the ambiguity surface of Eq. (8) corresponds to the contribution of the weak source field, and its maximization is expected to provide the location of the weak source. Again, the success of this method in estimating the location the weak source depends on the quality of $\mathbf{G}(\hat{r}_{str}, \hat{z}_{str})$ and $\hat{\alpha}_{str}$.

III. LOCALIZATION OF MULTIPLE SOURCES WITH MAXIMUM *a posteriori* ESTIMATION AND GIBBS SAMPLING

In Ref. 12 a Gibbs sampling scheme was presented developing an estimate of the joint posterior distribution $p(r, z_s, \alpha, \sigma^2 | \mathbf{X})$, where r and z_s are the source range and depth, respectively, α is the complex spectrum of the source, σ^2 is the variance of additive Gaussian noise, and $\mathbf{X}$ is the observed data at a set of L spatially (vertically here) separated hydrophones. From this joint distribution, the marginal posterior distribution $p(r, z_s | \mathbf{X})$ for source range and depth is calculated, maximization of which leads to MAP estimates for source location.[12,13] The estimate of the joint posterior distribution was derived using a Gibbs sampler[14–18,12] that draws samples in a cyclical fashion from marginal conditional posterior distributions.

The signal model in the frequency domain considered for the single source problem of Ref. 12 is as follows:

$$\mathbf{X} = \alpha \mathbf{G} + \mathbf{W}, \tag{9}$$

where $\mathbf{G}$ is a Green's function vector calculated at the L receiving phones, α is a complex factor containing source amplitude and phase, and $\mathbf{W}$ is an L-dimensional vector, each sample of which is drawn from a complex normal zero mean distribution with variance $2\sigma^2$.

The technique of Ref. 12 is extended in this paper for M sources, $M \geq 1$. The signal model then becomes

$$\mathbf{X} = \sum_{i=1}^{M} \alpha_i \mathbf{G}_i + \mathbf{W}. \tag{10}$$

Here, $\mathbf{G}_i = \mathbf{G}(r_i, z_{si})$, r_i and z_{si} are the range and depth of the ith source, α_i are complex factors containing amplitude and phase information for the ith source. The number of sources M is assumed known.

Using the same prior distributions as in Ref. 12, the joint posterior distribution of all parameters is as follows:

$$p(r_1, z_{s1}, \ldots, r_M, z_{sM}, \alpha_1, \ldots, \alpha_M, \sigma^2 | \mathbf{X})$$

$$\propto \frac{1}{(2\pi)^L} \frac{1}{\sigma^{2L+2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{X} - \sum_{i=1}^{M} \alpha_i \mathbf{G}_i(r_i, z_{si})\|^2\right).$$

(11)

For the marginal posterior distribution of α_i we need to fix r_j, z_{sj}, $j = 1, \ldots, M$, α_j, $j, = 1, \ldots, M$, $j \neq i$, and σ^2 in Eq. (11). The marginal posterior distribution of α_i conditional on all other unknowns can be rewritten as

$$p(\alpha_i | r_1, z_{s1}, \ldots, r_M, z_{sM}, \alpha_j (j = 1, \ldots, M; j \neq i), \sigma^2, \mathbf{X})$$

$$= M_{\alpha i} \exp\left(-\frac{\|\mathbf{G}_i\|^2}{2\sigma^2}\right.$$

$$\left. \times \|\left(\alpha_i - \left(\mathbf{G}_i^*\left(\mathbf{X} - \sum_{j=1,\ldots,M; j \neq i} \alpha_j \mathbf{G}_j\right)\right) / \|\mathbf{G}_i\|^2\right)\|^2\right),$$

(12)

where $M_{\alpha i}$ is a constant with respect to α_i. Complex variable α_i is, thus, normally distributed with mean $(\mathbf{G}_i^*(\mathbf{X} - \Sigma_{j=1,\ldots,M; j \neq i} \alpha_j \mathbf{G}_j))/\|\mathbf{G}_i\|^2$ and variance $2\sigma^2/\|\mathbf{G}_i\|^2$.

The conditional posterior distribution for the variance is identified as an inverse χ^2 distribution:

$$p(\sigma^2 | r_1, z_{s1}, \ldots, r_M, z_{sM}, \alpha_1, \ldots, \alpha_M, \mathbf{X})$$

$$\propto \frac{1}{\sigma^{2L+2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{X} - \sum_{i=1}^{M} \alpha_i \mathbf{G}_i(r, z_s)\|^2\right).$$

(13)

Samples from both the normal and χ^2 distributions can be readily generated.[19]

The conditional distribution of r_i and z_{si} is as follows:

$$p(r_i, z_{si} | r_j, z_{sj} (j = 1, \ldots, M; j \neq i), \alpha_i (i = 1, \ldots, M), \sigma^2, \mathbf{X})$$

$$= K \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{X} - \sum_{j=1}^{M} \alpha_j \mathbf{G}_j(r_j, z_{sj})\|^2\right),$$

(14)

where K is a constant. In the distribution of Eq. (14), r_j and z_{sj} for all j are fixed with the exception of r_i and z_{si}, $i \neq j$. The distribution of Eq. (14) is evaluated on a grid for r_i and z_{si} similar to the implementation in Ref. 20 and following the griddy Gibbs modeling of Ref. 21, because no closed form is known.

Gibbs sampling begins with a set of randomly chosen initial conditions for all unknown parameters (r_i, z_{si}, α_i, and σ^2, $i = 1, \ldots, M$). The process as implemented here first draws a sample from the conditional distribution of range r_1 and source depth z_{s1}; this two-dimensional sample contains the new, updated values of range and depth for the first source and first iteration. Subsequently, a sample is drawn from the Gaussian marginal conditional posterior of α_1 [Eq. (12)]. The same process is repeated for r_j, z_{sj}, α_j, j
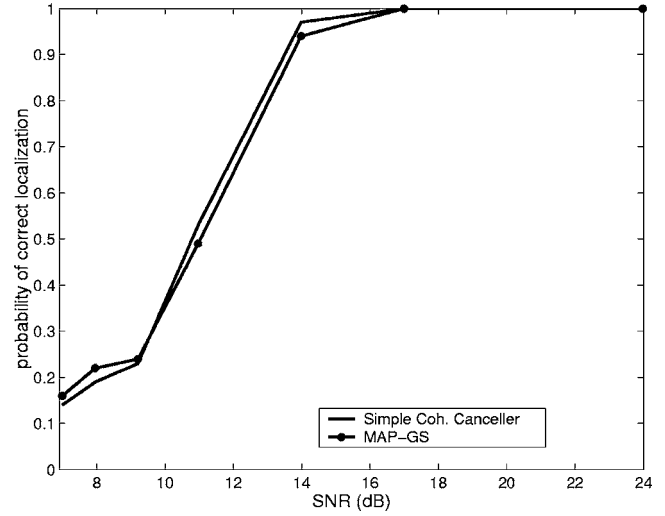
FIG. 1. Probability of correct localization vs SNR for 600 Hz for the simple coherent Bartlett canceller and the GS-MAP processor.

$= 2, \ldots, M$. Lastly, a sample is drawn for the variance from the inverse χ^2 distribution of Eq. (13), completing the first iteration. For a large number of iterations, the obtained sample sequences eventually converge to the true joint posterior distribution of r_i, z_{si}, α_i, $i = 1, \ldots, M$, and σ^2.[17,14,15]

IV. PERFORMANCE EVALUATION OF THE MAXIMUM *a posteriori* PROCESSOR

To evaluate the proposed method, we generated synthetic data simulating sound propagating in a shallow water environment. Received field data at a vertical array of 24 phones were generated corresponding to two sound sources. A strong source was at a range and depth of 4 km and 2 m, respectively. A weaker source was at a range and depth of 2.4 km and 48 m. The complex amplitudes of the strong source and weak source were $\alpha_1 = \alpha_{str} = 20 \exp(i/2)$ and $\alpha_2 = \alpha_w = 2 \exp(3i/2)$.

SNR varied between 7 and 24 dB. For the performance evaluation we generated 100 noisy realizations per case. We then performed source localization with a simple incoherent cancellation Bartlett processor [Eq. (7)], a simple coherent cancellation Bartlett processor [Eq. (8)], and the MAP estimator proposed here, and, for each method, estimated probabilities of correct localization. Correct localization is here defined as obtaining estimates between 2.22 and 2.58 km for range and between 42 and 54 m for depth for the weak source and between 3.82 and 4.18 km for range and 0 and 8 m for depth for the strong source. The search interval for range was between 0 and 5 km with a 2 m spacing; the interval for depth was between 0 and 72 m with a 2 m spacing. The Gibbs sampler was run for 2000 iterations, the first 200 of which were discarded. Initial values used in the Gibbs sampler were 10 for σ^2, 100 for complex amplitudes of both sources, 2 km and 18 m for range and depth for the weak source, and 3 km and 18 m for range and depth for the strong source. Experimentation was performed with higher iteration numbers and different initial conditions without significant changes in the results.

Figure 1 shows probability of correct localization for the
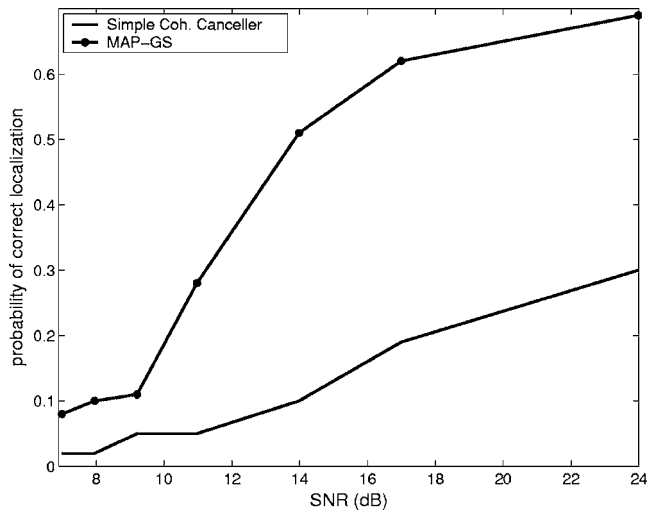
FIG. 2. Probability of correct localization vs SVR for 200 Hz for the simple coherent Bartlett canceller and the GS-MAP processor.

FIG. 4. Posterior probability distribution of α_{str}, the complex amplitude of the strong source.

coherent cancellation Bartlett processor and the MAP processor, where both sources transmit at 600 Hz; probability of correct localization as obtained for the incoherent cancellation Bartlett processor was low and is not included in the figure. The two processors perform similarly, with the simple processor being slightly better at high SNRs and the MAP processor optimized via Gibbs sampling being minimally superior at lower SNRs.

Figure 2 shows probability of correct localization versus SNR for the simple coherent Bartlett canceller and the MAP processor for a frequency of 200 Hz; in this case as well, the incoherent canceller provided low probability of correct localization and is not included in the figure. The performance of the MAP processor dominates that of the coherent Bartlett processor for the entire SNR range.

An explanation for the poor performance of the coherent Bartlett-CLEAN processor at this frequency can be found in Fig. 3, which shows probability of source localization for the coherent Bartlett-CLEAN and MAP processors (also shown

FIG. 3. Probability of correct localization vs SVR for 200 Hz for the simple coherent Bartlett canceller, the GS-MAP processor, and a coherent Bartlett canceller that knows exactly the complex amplitude of the strong source.

in Fig. 2) and the probability of correct localization for the coherent Bartlett-CLEAN processor that employs as $\hat{\alpha}_{str}$ the true value of α_{str} [it does not assume $r_{s,str}$ and $z_{s,str}$ to be known, though, and estimates them by maximizing the ambiguity surface of Eq. (1)]. The latter processor is not realistic because information on α_{str} is not typically available but offers an upper performance bound for the coherent Bartlett-CLEAN processor. For high SNRs, this ideal processor outperforms both the MAP and the realistic coherent Bartlett processor. As the SNR decreases, the MAP processor becomes slightly better than the "artificial" processor. The simple coherent processor always lags behind substantially. Figure 3 illustrates that the cause of the poor performance of the simple coherent Bartlett-CLEAN processor lies in the estimation of α_{str}, which is necessary before Eq. (8) is used.

Figure 4 demonstrates why the MAP processor does not suffer greatly from the same problem. The figure shows the posterior probability distribution of the complex amplitude of the strong source as obtained from the MAP processor. There is significant probability concentrated at a phase of 0.5 rad (true value of the phase of that source). There is more uncertainty regarding the source amplitude; we see that the probability distribution exhibits high values for amplitudes roughly between 7 and 24 (the true amplitude is 20). The plot also indicates the point estimate of α_{str} as obtained from the coherent Bartlett-CLEAN processor, which is far away from the true amplitude and phase of the source (it coincides with a secondary mode of the probability distribution). The coherent Bartlett-CLEAN estimator fails to estimate the weak source location using that point estimate. The MAP processor, integrating over the probability distribution of the strong source complex amplitude, correctly identifies the weak source.

More simulations were run with data at frequencies between 200 and 600 Hz. It was observed that for these frequencies the two processors had similar performance. With frequencies at and below 200 Hz, there is a significant advantage in using the proposed MAP-Gibbs sampling processor. Figure 5 shows probability of correct localization as ob-
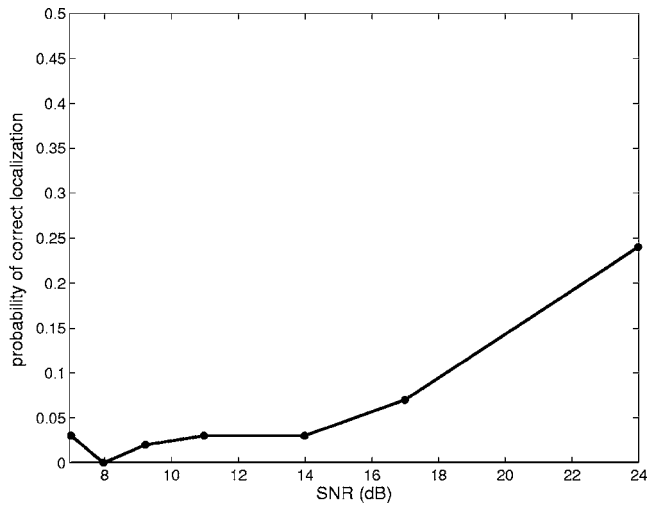
FIG. 5. Probability of correct localization vs SNR for 100 Hz for the GS-MAP processor.

tained from the new processor for the two sources operating at 100 Hz. Although these probabilities are not particularly high, the processor performs better than the simple coherent processor (not shown here) which fails to successfully localize both sources in every single case.

Observing more closely the MAP-Gibbs sampling coherent cancellation results, we observed that, although even in high SNRs the processor failed in several cases to estimate the two source locations correctly, the approach was, surprisingly, quite successful in depth estimation. Figure 6 shows probability of correct depth estimation for both the simple and the proposed processor. The simple coherent processor has a probability of correct depth estimation between 0.33 and 0.45 for all SNRs with an arbitrary pattern of variations. The proposed processor has a probability of correct depth estimation that increases as SNR increases (as expected) and that reaches 0.88 for a SNR of 24 dB. The ability to identify correctly source depth in an environment with interference
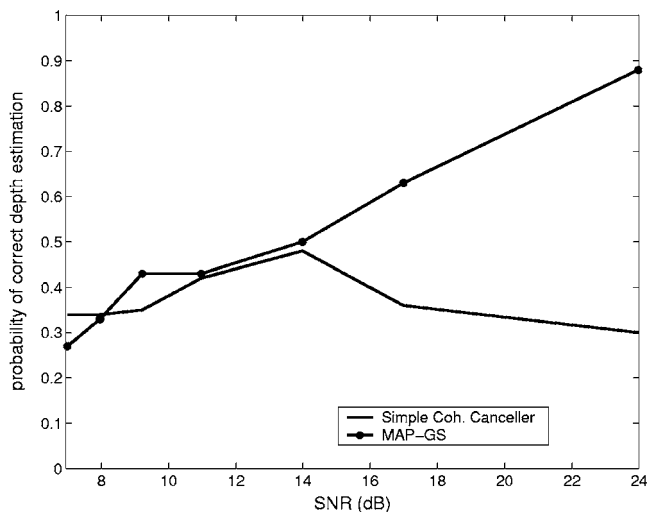
FIG. 6. Probability of correct depth estimation vs SNR for 100 Hz for the simple coherent Bartlett canceller and the GS-MAP processor.

present is significant, because it answers the question of whether a sensed source is submerged or whether it is located at the surface of the water column.

A coherent multiple source localization approach based on MFP was also proposed in Ref. 6. There, the received signal was modeled as follows:

$$\mathbf{X} = \sum_{i=1}^{M} b_i \mathbf{G}_i(r_i, z_{s,i})/\|\mathbf{G}_i(r_i, z_{s,i})\| + \mathbf{W}, \qquad (15)$$

where $\mathbf{G}$ is a Green's function vector calculated at the L receiving phones and $\mathbf{W}$ is an L-dimensional complex, zero-mean noise vector. The model of Eq. (15) combines contributions from M sources without involving source levels, which are actually estimated by our method. Coefficients b_i represent received levels: $|b_i|$ is the strength of the contribution of the ith source at the receiving array. As mentioned in Ref. 6, source levels were avoided in the modeling in order to circumvent correlations between them and source ranges which might potentially complicate the estimation process. Using the model of Eq. (15) and simulated annealing, a search was performed in Ref. 6 for the number of sources, their location, and environmental parameters.

To compare our processor to that of Ref. 6, we performed source localization of two sources using the model of Eq. (15) and estimating posterior distribution $p(r_1, z_{s1}, r_2, z_{s2}, b_1, b_2, \sigma^2 | \mathbf{X})$ using Gibbs sampling. In some cases the results of the process relying on Eq. (15) were practically identical to the results of the processor developed here [which uses the model of Eq. (9)]. In other cases, however, after the strong source was identified, the sampling process was "trapped" in this source's neighboring region; as a result, the weak source was not identified and a second source ("shadow" of the actual strong source) was falsely found. Equation (15) removes the effect of two very different source amplitudes from the data; this "smoothing" might account for the difficulty of the processor relying on Eq. (15) in identifying both sources. The problem can be remedied by applying restrictions prohibiting the Gibbs sampler from searching for the weak source in the vicinity of the strong source. Such restrictions are not necessary for the processor proposed in this work.

V. UNKNOWN NUMBER OF SOURCES

In the preceding section, the number of sources was assumed known and set equal to two. In practice there is no such precise information on the number of sources. In this section, the assumption of a known source count is relaxed; along with complex source amplitudes, locations, and noise variance, the number of sources is estimated as well.

According to the analysis presented in this paper, the joint probability distribution we have so far estimated with Gibbs sampling is in reality a conditional distribution, conditioning being on the number of sources M: $p(\alpha_1, \ldots, \alpha_M, r_1, z_{s1}, \ldots, r_M, z_{sM}, \sigma^2 | M, \mathbf{X})$. Following the Bayesian paradigm, a prior distribution can be specified for M, and, in like manner with the other parameters, a posterior distribution can be estimated from which we can infer the number of sources present.
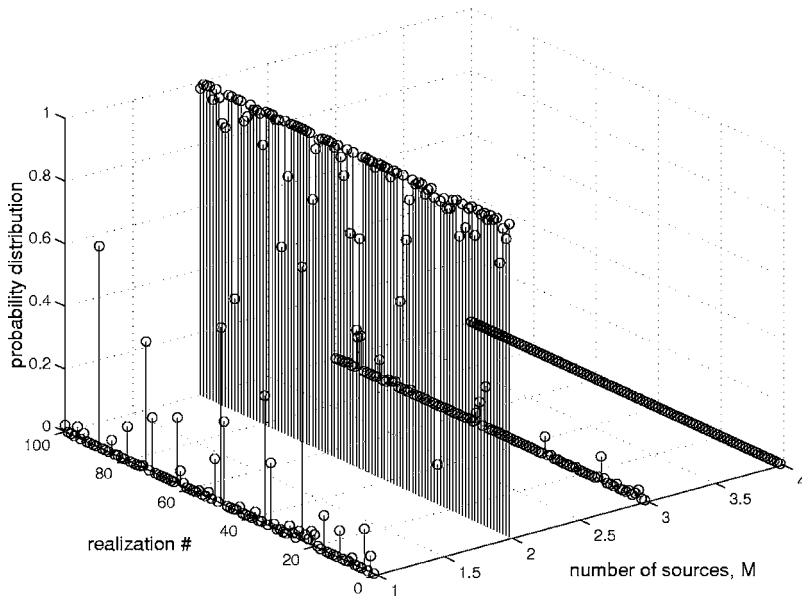
FIG. 7. Posterior distribution of M vs M and realization number: SNR is 14 dB.

In the absence of specific information on M, we select a uniform prior over a range believed to be realistic:

$$p(M) = \frac{1}{M_2 - M_1 + 1}, \quad M_1 \leq M \leq M_2, \tag{16}$$

where M_1 and M_2 are lower and upper bounds for the expected number of sources.

Having already calculated the joint posterior probability distribution of all unknowns, we then compute the marginal posterior distribution $p(M|\mathbf{X})$, where

$$p(M|\mathbf{X}) = \int_{\sigma^2} \int_{\alpha_M} \cdots \int_{\alpha_1} \int_{r_M} \int_{z_{sM}} \cdots \int_{r_1} \int_{z_{s1}}$$

$$p(\alpha_1, \ldots, \alpha_M, r_1, z_{s1}, \ldots, r_M, z_{sM}, \sigma^2 | M, \mathbf{X}) p(M)$$

$$d_{\alpha_1} \cdots d_{\alpha_M} d_{r_1} dz_{s1} \cdots dr_M dz_{sM} d\sigma^2. \tag{17}$$

Maximizing Eq. (17) yields a MAP estimate of M.

One hundred noisy data realizations were generated to test the estimation of the number of sources; the SNR was 14 dB and the frequency was 200 Hz. Two sources were present as in the previous section. Using prior knowledge, it was assumed that M could vary between 1 and 4. The estimated marginal posterior distributions of M for all realizations are exhibited in Fig. 7. It is evident that most distributions are maximized for $M=2$, the true number of sources. Specifically, 99 out of 100 times the maximum occurs for $M=2$; for one realization, the posterior distribution is maximized for $M=1$.

In addition to the point estimate $\hat{M}=2$, our approach also provides information on the uncertainty in the estimation of the number of sources. From the distribution of Fig. 7, we can see that there is a very small probability of three or four sources being present. It appears that the number of sources is probably two, with one source being the next best choice.
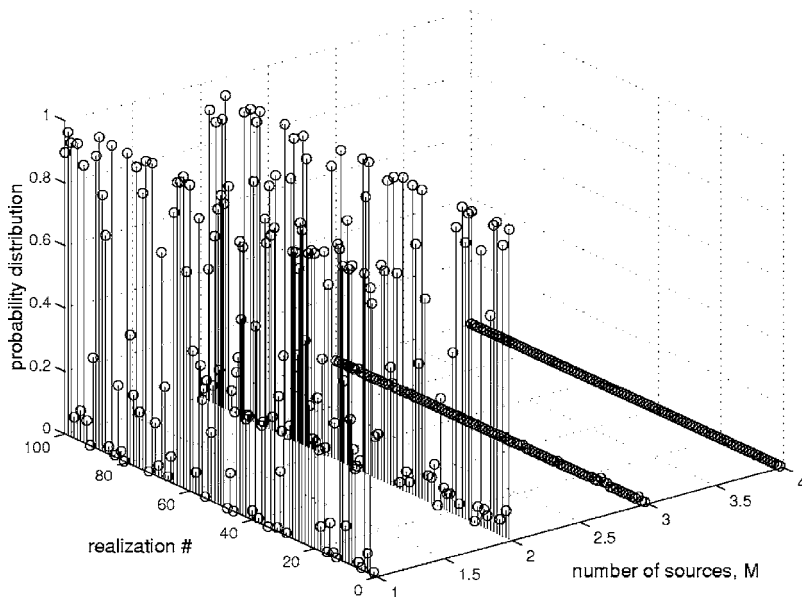
FIG. 8. Posterior distribution of M vs M and realization number: SNR is 11 dB.
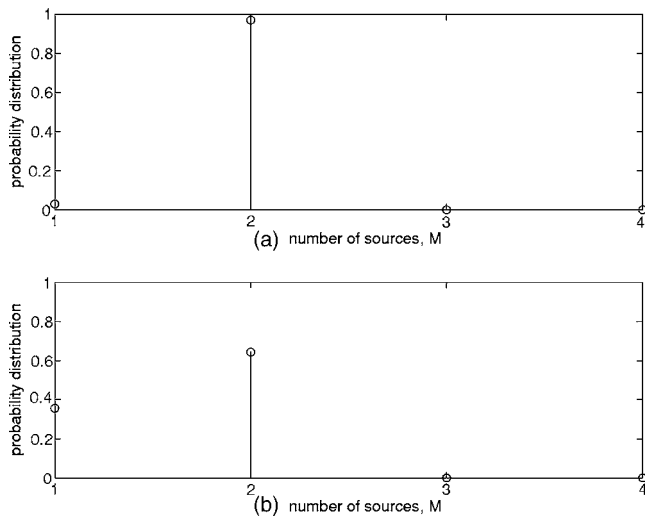
FIG. 9. Posterior distribution of M vs M for a single data realization: SNR is (a) 14 dB and (b) 11 dB.

As expected, results deteriorate as the noise increases. Figure 8 shows the marginal posterior distributions of M for all realizations for a SNR of 11 dB. A MAP estimate of 1 is obtained for 50% of the realizations. For the remaining 50%, an estimate $\hat{M}=2$ is obtained, representing the true number of sources.

Figure 9 shows the posterior distribution of M for a *single* data realization for each of the two considered SNRs. The distribution of M for the lower SNR [Fig. 9(b)] exhibits higher variability than the distribution for the higher SNR [Fig. 9(a)] with the probability divided between values 2 and 1 for M because of the increased noise level.

VI. PERFORMANCE EVALUATION OF THE MAP PROCESSOR IN THE PRESENCE OF MISMATCH

Performance evaluation in Sec. IV was performed assuming perfect knowledge of the propagation environment. This assumption is now relaxed and mismatch is introduced in the water column sound speed profile. Figure 10 shows the water column sound speed profile employed both for data
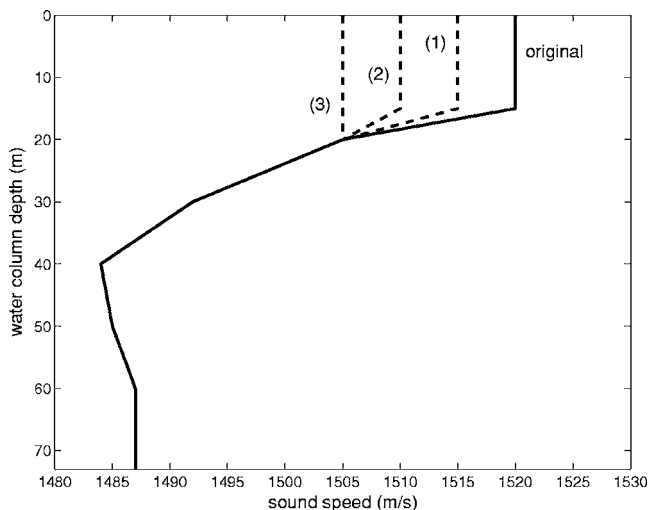
FIG. 10. Sound speed profiles used for data and replica generation.

TABLE I. Probability of correct localization for the strong source for the new MAP-Gibbs sampling and the simple processors for no mismatch and three mismatched sound speed profiles.

Processor	No mis.	Prof. 1	Prof. 2	Prof. 3
MAP-GS	1.00	0.60	0.30	0.24
Simple coherent canc.	1.00	0.18	0.03	0.00

and replica generation in Sec. IV. In the same figure, new sound speed profiles are presented [dashed lines, identified as (1)–(3)] that are used in the present section for data generation; replica generation is still performed with the original, solid-line sound speed profile. The new MAP processor and the simple incoherent Bartlett canceller were then applied to 100 noisy data realizations generated for a frequency of 200 Hz with the three new sound speed profiles. A strong source was again at a range and depth of 4 km and 2 m, respectively, and a weaker source was at a range and depth of 2.4 km and 48 m. The SNR was 17 dB.

Source localization results from both processors deteriorate dramatically even for the smallest mismatch considered here [profile (1) in Fig. 10], with the weak source becoming nonidentifiable. It is of interest, however, that the new processor localizes the strong source at a much better rate than the simple Bartlett canceller. Table I shows probabilities of correct localization for the strong source for four cases: no mismatch, and mismatch as introduced by sound speed profiles (1)–(3). With no mismatch between assumed and true sound speed profiles, the strong source is perfectly localized with both processors. The sound speed mismatch causes the simple coherent cancellation performance to drop very quickly. The MAP-Gibbs sampling performance, although affected by the mismatch, is considerably better than that of the simple processor.

VII. CONCLUSIONS

A MAP matched-field estimation approach was developed for localization of multiple sources. The method, optimized with Gibbs sampling, provides estimates of the number of sources and their locations. The method is, in essence, a deconvolution approach that can remove strong interfering source contributions from the received field for the identification of weaker sources of interest. The new approach was evaluated against simple interference cancellation methods for multiple source localization and was found superior for frequencies around and below 200 Hz. The superiority of the method is attributed to integrating over uncertainty instead of relying on point estimates as other methods do. Unlike previously developed methods, our approach also offers posterior probability distributions of the parameters of interest demonstrating the uncertainty present in the estimation problem. With the marginal conditional distributions of source levels and phases readily available, the approach can also be employed for source signature estimation. The method as presented here does not integrate over environmental uncertainty. Simulations indicate that, even without such integration, the processor is more robust with respect to mismatch than simple coherent cancellation. The proposed method can

be further extended by considering environmental uncertainty; estimates of environmental parameters can be further obtained as in Ref. 18.

ACKNOWLEDGMENTS

[1] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993).

[2] A. Baggeroer, W. Kuperman, and P. N. Michalevsky, "An overview of matched field processing in ocean acoustics," IEEE J. Ocean. Eng. **18** (4), 401–424 (1993).

[3] J. Ozard, "Matched field processing in shallow water for range, depth, and bearing determination: Results of experiment and simulation," J. Acoust. Soc. Am. **86** (2), 774–753 (1989).

[4] A. N. Mirkin and L. H. Sibul, "Maximum likelihood estimation of the locations of multiple sources in an acoustic waveguide," J. Acoust. Soc. Am. **95**, 877–888 (1994).

[5] M. D. Collins, L. T. Fialkowski, W. A. Kuperman, and J. S. Perkins, "The multi-valued Bartlett procesor and source tracking," J. Acoust. Soc. Am. **97**, 235–241 (1995).

[6] M. V. Greening, P. Zakarauskas, and S. E. Dosso, "Matched-field localization for multiple sources in an uncertain environment with application to Arctic ambient noise," J. Acoust. Soc. Am. **101**, 3525–3538 (1997).

[7] L. Zurk, N. Lee, and J. Ward, "Source motion mitigation for adaptive matched field processing," J. Acoust. Soc. Am. **113** (5), 2719–2731 (2003).

[8] T. B. Neilsen, "Localization of multiple acoustic sources in the shallow ocean," J. Acoust. Soc. Am. **118**, 2944–2953 (2005).

[9] H. C. Song, J. de Rosny, and W. A. Kuperman, "Improvement in matched field processing using the CLEAN algorithm," J. Acoust. Soc. Am. **113**, 1379–1386 (2003).

[10] J. Hőgbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," Astron. Astrophys., Suppl. Ser. **15**, 417–426 (1974).

[11] U. J. Schwarz, "Mathematical-statistical description of the iterative beam removing technique (method CLEAN)," Astron. Astrophys. **65**, 345–356 (1978).

[12] Z.-H. Michalopoulou, "The effect of source amplitude and phase on matched field source localization," J. Acoust. Soc. Am. **119**, EL21–EL26 (2006).

[13] A. Richardson and L. Nolte, "A posteriori probability source localization in an uncertain sound speed, deep ocean environment," J. Acoust. Soc. Am. **89** (5), 2280–2284 (1991).

[14] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721–741 (1984).

[15] A. E. Gelfand and A. F. Smith, "Sampling-based approaches to calculating marginal densities," J. Am. Stat. Assoc. **85**, 398–409 (1990).

[16] J. Besag and P. Green, "Spatial statistics and Bayesian computation," J. R. Stat. Soc. Ser. B (Methodol.) **55** (1), 25–37 (1993).

[17] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, 1st ed. (Chapman and Hall, CRC, 1996).

[18] S. Dosso, "Quantifying uncertainty in matched field inversion. I A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111** (1), 129–142 (2002).

[19] L. Devroye, *Non-Uniform Random Variate Generation* (Springer, Berlin, 1986).

[20] Z.-H. Michalopoulou and M. Picarelli, "Gibbs sampling for time delay and amplitude estimation in underwater acoustics," J. Acoust. Soc. Am. **117**, 799–808 (2005).

[21] M. A. Tanner, *Lecture Notes in Statistics: Tools for Statistical Inference*, Observed Data and Data Augmentation Methods Vol. **67**, 2nd ed. (Springer, Berlin, 1992).

Cluster analysis and robust use of full-field models for sonar beamforming

Brian Tracey,[a] Nigel Lee, and Srinivas Turaga[b]

MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420

Multipath propagation in shallow water can lead to mismatch losses when single-path replicas are used for horizontal array beamforming. Matched field processing (MFP) seeks to remedy this by using full-field acoustic propagation models to predict the multipath arrival structure. Ideally, MFP can give source localization in range and depth as well as detection gains, but robustly estimating range and depth is difficult in practice. The approach described here seeks to collapse full-field replica outputs to bearing, which is robustly estimated, while retaining any signal gains provided by the full-field model. Cluster analysis is used to group together full-field replicas with similar responses. This yields a less redundant "sampled field" describing a set of representative multipath structures for each bearing. A detection algorithm is introduced that uses clustering to collapse beamformer outputs to bearing such that signal gains are retained while increases in the noise floor are minimized. Horizontal array data from SWELLEX-96 are used to demonstrate the detection benefits of sampled field as compared to single-path beamforming. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2346128]

PACS number(s): 43.30.Wi, 43.60.Mn, 43.60.Jn [EJS]   Pages: 2635–2647

I. INTRODUCTION

Horizontal array beamforming for passive sonar is often based on the assumption that the signals impinging on the array arrive via a single propagation path. A model accounting for wave-front curvature across the array may be used near broadside, while a planewave model is used near endfire or for distant sources. However, multipath acoustic propagation is often important in shallow water. The use of a single-path model can therefore introduce a mismatch that reduces the output power of targets and may significantly limit the effectiveness of adaptive interference cancellation techniques.[1] The effects of multipath propagation are most noticeable at endfire, as the ability of horizontal arrays to resolve multipath is best for sources in this sector.[2] The approach described in this paper seeks to improve horizontal array performance near endfire in the presence of multipath.

Matched field processing (MFP) has been investigated as a means for accounting for multipath.[3] MFP uses full-field acoustic models to calculate replica vectors that represent the array response for a three-dimensional grid of possible source locations. Typically, MFP attempts to achieve simultaneous improvements over single-path beamformers in both detection performance (by using physics-based signal models to reduce signal mismatch) and source localization (by providing range and depth in addition to bearing).

Experimental use of MFP has shown that robust range and depth localization is difficult to achieve. In many cases this is due to uncertainty in the environmental inputs required by acoustic propagation models. Robustness can be increased by improving environmental calibration to give a more accurate set of replicas[4–6] or by explicitly accounting for uncertainty.[7–10] Another problem with MFP, especially for horizontal arrays, is that the array response may be very similar for sources at widely separated ranges and depths. Range and depth estimation therefore involves choosing the correct source location from a set of highly ambiguous beamformer outputs, which requires very accurate calibration and high signal-to-noise ratio.[11]

The signal processing approach described below is motivated by two observations. First, simulations described in Sec. III were used to investigate MFP localization accuracy for horizontal arrays when environmental information is uncertain. The results indicate that localization in bearing is much more tolerant of environmental uncertainty than is localization in range and depth. Second, experimental results in Sec. V demonstrate that use of MFP replicas can lead to reduced mismatch losses and therefore potentially better detection. These two observations lead to the following question: Is it possible to use MFP replicas in a way that achieves mismatch reduction, *without* attempting range and depth localization? Answering this question requires a method for collapsing the three-dimensional (3D) MFP beamformer output to bearing.

A simple approach is to maximize over MFP beamformer outputs for all ranges and depths along each bearing. This preserves any gains on signal provided by MFP, but leads to a significant increase in the displayed noise floor. As shown below, the resulting output shows little improvement in signal detectability.

An alternate procedure is developed here that includes a detection stage before collapsing the beamformer output to bearing. The procedure begins by analyzing the set of predicted MFP replicas and representing them by a set of *clusters*.[12] As described in Sec. II, each cluster corresponds to a set of spatial locations that have essentially the same

---

[a] Current address: Neurometrix, Inc. 62 Fourth St. Waltham, MA.
[b] Current address: Department of Brain and Cognitive Science, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA.

multipath structure (i.e., replicas that are highly correlated with each other). In standard MFP these locations would appear as highly ambiguous regions in the output. The cluster analysis output is denoted here a "sampled field," as the number of clusters, or distinct multipath structures, is often much smaller than the number of possible three-dimensional source locations (from this point of view, the full set of MFP replicas can be considered an "over-sampled field," as it contains redundant information). Because the set of clusters captures the diversity of multipath predicted by the full-field model, any mismatch reductions provided by the full-field model are retained.

After cluster analysis is complete, beamforming is done using a representative replica vector from each cluster (this gives computational gains relative to MFP,[13] but computational gain is not the primary objective of the current work). The cluster outputs are then analyzed. If a cluster has higher output power than other clusters that are highly correlated with it, it is judged to be signal-like. Otherwise, the cluster is judged to represent either noise or sidelobe energy from a nearby source, and its output power is set to an estimate of the noise level. After this detection stage is complete, the data are collapsed to the bearing dimension. Used in this manner, clustering allows an automated analysis of ambiguous regions in the output so that decisions can be made as to which set of search locations may contain the signal. Similar strategies have been used in analysis of radar outputs.[14] Thus any signal gains in the proposed approach result from the use of full-field replicas, while clustering is used to collapse the beamformer outputs to a bearings-only display.

An alternative method for collapsing MFP output to a bearings-only display has been proposed by Yang and Yates.[15] Their work requires an iterative search over range and depth to refine estimates of source bearing, and therefore it implicitly requires 3D localization. By contrast, the sampled field approach presented here collapses full-field outputs to bearings only without requiring range and depth localization. In cases where the environmental knowledge is good, matched field range and depth localization could be attempted as a follow-on stage to the sampled field bearing localization described here.

Section II below describes cluster analysis in more detail, while Sec. III outlines simulations used to study localization errors due to environmental uncertainty. The cluster-based signal processing approach is detailed in Sec. IV and is labeled sampled field processing (SFP) to distinguish it from standard MFP. Section V analyzes data from the SWellEx-96 experiment, showing that the approach improves signal detectability at endfire as compared to planewave or MFP approaches, especially when adaptive beamforming is employed.

II. CLUSTER ANALYSIS OF FULL-FIELD REPLICAS

Full-field MFP replicas are based on the Green's function, or channel response, for a set of possible source locations in range, depth, and azimuth. The results below use a range-independent normal mode model to calculate the Green's function,[16] although clustering can be applied to the

output of any ocean propagation model. Taking $(x, y) = (0, 0)$ to be at the center of the array, with $z = 0$ at the ocean surface, the Green's function for a sensor at $\mathbf{r}_i = (x_i, y_i, z_i)$ is given by

$$G(f, \mathbf{r}_i, \Theta) = \frac{ie^{-i\pi/4}}{\rho(z_s)\sqrt{8\pi}} \sum_{n=1}^{N} \psi_n(z_s)\psi_n(z_i)\frac{e^{-ik_n r_i}}{\sqrt{k_n}}, \quad (1)$$

where ψ and k_n are the mode shapes and mode horizontal wave numbers at the frequency f, z_s is the source depth, $r_i = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2}$ is the source-receiver range, and $\Theta = (x_s, y_s, z_s)$ is the source location in three dimensions.

The full-field replica vectors are computed as normalized versions of the Green's function:

$$\mathbf{v}(f, \Theta) = \frac{\mathbf{G}}{|\mathbf{G}|}\sqrt{N}, \quad (2)$$

where $\mathbf{G}$ is a vector of calculated Green's functions for each receiver. By contrast, planewave or range-focued replicas are calculated using a single-path propagation model

$$\mathbf{v}_{pw}(f, \Theta) = e^{-ik\mathbf{r}(\Theta)}, \quad (3)$$

where $k = \omega/c$ is the wave number for a wave traveling at speed c, and $\mathbf{r}(\Theta)$ is a vector of ranges between each array element and the hypothesized source location.

The set of MFP replica vectors can be analyzed using clustering, a technique that automatically breaks a set of observations into *clusters*, or groups of vectors that are similar to each other.[12] This similarity is defined by a distance metric specified by the user. For each cluster, a list of cluster members is returned along with a cluster center. The cluster center can be used as a representative of all the members of the cluster. The clustering algorithms used here require the user to specify a cluster *radius*, which is the maximum distance allowed between any point in a cluster and its center.

When applying clustering to beamforming, a natural distance metric is provided by the mismatch between pairs of replicas. The measure has been used by Cox[17] and is defined as

$$\text{dist}(\mathbf{v}_i, \mathbf{v}_j) \triangleq 1 - \frac{\|\mathbf{v}_i^H \cdot \mathbf{v}_j\|^2}{\|\mathbf{v}_i\|^2 \|\mathbf{v}_j\|^2}, \quad (4)$$

where the frequency domain replica vectors are given by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. Replica pairs that are highly correlated will have low mismatch and will tend to be clustered together.

Using this distance metric, a cluster is defined as a set of replicas $\mathbf{v}_i$ that all have a distance from a cluster *center* that is less than some threshold η

$$\text{dist}(\mathbf{v}_i, \tilde{\mathbf{v}}_j^{\text{cen}}) \leq \eta. \quad (5)$$

The cluster center $\tilde{\mathbf{v}}_j^{\text{cen}}$ is chosen from the original set of full-field replicas.

Clustering of MFP replicas can be computationally intensive and requires the use of efficient algorithms. The method used here is a "greedy" algorithm that first computes a matrix containing the mismatch between every pair of vectors. The vector with the largest number of close matches (i.e., replicas with mismatch less than the threshold η) is picked as the first cluster center. All replica vectors closer
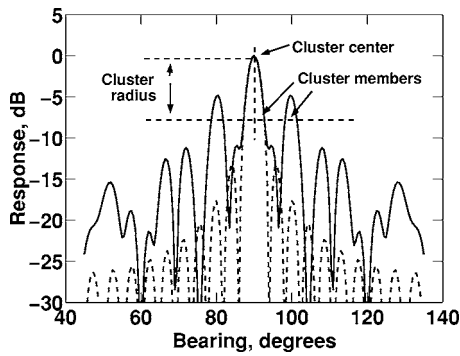
FIG. 1. Beam patterns for nominal well-sampled array (dashed line) vs sparse array (solid line). For a sparse array, steering directions that are highly correlated with the main axis of response may be widely separated in space. Clustering provides a way to identify these regions of similar response.

FIG. 2. Plot showing members of an example cluster formed from MFP replicas for a single-line array at 94 Hz. The cluster (with 1.5 dB cluster radius) is well contained in azimuth and does not spread over many azimuthal beams. For a given azimuth, high correlation between many range and depth locations is evident.

than the threshold are assigned to this center. The cluster center and its cluster members are removed from the full set of replicas, and the process is repeated until all replicas are clustered. A more complete description of clustering algorithms is given in the Appendix.

The application of clustering to beamforming can be made clearer by considering a planewave example. Figure 1 compares beampatterns for well-sampled and sparse line arrays steered to a bearing of 90°. The steering angle is equivalent to a cluster center. The beam extent is defined in terms of a maximum allowable loss with respect to the steering angle; this maximum allowable loss is the equivalent of the cluster radius. When high ambiguities are present, as occurs with sparse arrays, beams that are within the specified cluster radius may be spatially disjoint. Clustering provides an automated way to identify these disjoint regions of similar response.

The following simulation example shows how cluster analysis can be applied to MFP replicas. Simulations were performed for a 101 element, 200 m long horizontal line array situated at a depth of 120 m in the Santa Barbara Channel environment[18] (see Table I for details). Full-field MFP replicas were generated for a bearing sector around endfire, for depths across the full water column and for ranges up to 10 km.

A typical cluster generated from the set of MFP replicas is shown in Fig. 2. This plot shows the locations of all MFP replicas that are identified as belonging to the same cluster. The cluster radius in decibels ($10 \log_{10} \eta^2$) is 1.5 dB for this example, meaning that each replica included in the cluster has at most 1.5 dB mismatch with the cluster center.

From Fig. 2 one can observe some characteristics that are generally true of MFP replicas for horizontal arrays. First, bearing is much better resolved than range or depth. This is a result of the array geometry, which has good sampling of azimuthal arrivals but poor sampling of the multipath structure used to infer range and depth. Range and depth estimation therefore requires finding the peak of a set of points that are very highly correlated. Second, although the cluster shown in Fig. 2 extends over many range and depth cells, there are many gaps in its coverage along the bearing shown. Multiple replicas (or clusters) are needed to provide low-mismatch coverage for all source locations along a given azimuth.

The 1.5 dB cluster radius for Fig. 2 was chosen based on beam spacings often used in planewave beamforming. Decisions about how widely to space steering vectors are commonly made by setting a maximum allowable loss relative to the beam's steering angle. While beams can be overlapped at the −3 dB point (corresponding to a 3 dB cluster radius), they may be spaced more finely to reduce loss on targets that lie between beams.

TABLE I. Parameters for sound speed and geoacoustic model used in simulations: z=depth from surface; c_c =compressional sound speed; ρ=density; α_c=compressional wave attenuation. The water column extends to 209 m depth and is followed by a layered seabed.

z (m)	c_c (m/s)	ρ (g/cm^3)	α_c (dB/λ)
0	1500	1	0
25	1500	1	0
75	1494	1	0
209	1488	1	0
209	1607	1.95	0.37
309	1702	1.95	0.37
309	1862	1.98	0.035
609	2374	1.98	0.035
609	2374	2.03	0.04
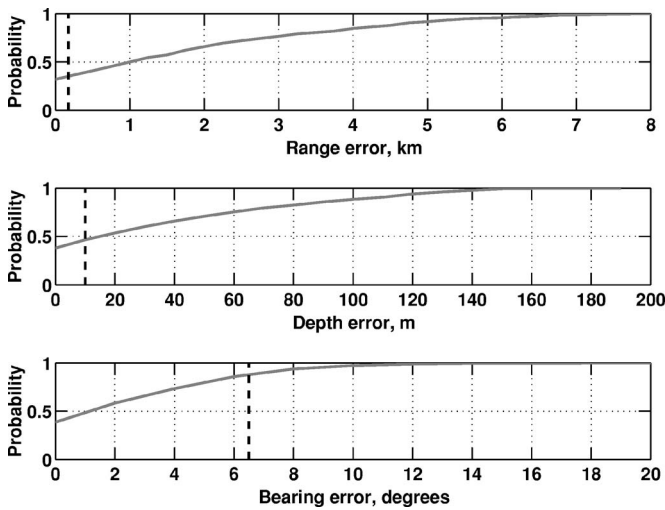
FIG. 3. Cumulative distribution functions for localization errors, single line HLA. Mainlobe widths for range, depth, and bearing are shown as dashed lines on each plot. Bearing errors are typically within the mainlobe, while median range and depth errors exceed the mainlobe size.

III. ARGUMENT FOR BEARINGS-ONLY LOCALIZATION

As will be discussed in Sec. IV, clustering enables efficient collapsing of MFP outputs to bearings-only localization. This section motivates why bearings-only localization using MFP replicas is of interest. Ambiguity surface patterns such as those shown in Fig. 2 suggest that bearing estimates found using MFP-derived replicas should be more robust than range or depth estimates. Monte Carlo studies of the effects of environmental mismatch are used here to verify this intuition.

Simulations were carried out for the line array and environment described above. Simulated data was generated for 200 random source locations, all at endfire bearing to the array but with ranges and depths uniformly distributed over the search space (0–200 m depth and 2–10 km range). No noise was added. MFP replica vectors were generated using mismatched environmental models, in which the initial compressional speed in the sediment was perturbed. Perturbations of ±10 and ±20 m/s were used for a series of runs whose results were accumulated. These mismatched replicas were used in beamforming. The 3D location of the peak beamformer output was found for each random source location.

Figure 3 shows the cumulative distribution function (CDF) of MFP localization errors caused by environmental uncertainty. This result shows that bearing errors were generally low, with the median error being roughly 1/3 of the beamwidth. In contrast, the median error in the range and depth dimensions exceeded the estimated mainlobe resolutions in these dimensions, which were estimated as described in.[18] This underscores the difficulty of the 3D localization problem.

Signal-to-noise ratio (SNR) is another important consideration for MFP localization.[11] As suggested by Fig. 2, many locations in the range/depth plane have very similar response. Estimating the correct range and depth from this flat parameter space therefore requires higher SNR than estimat-
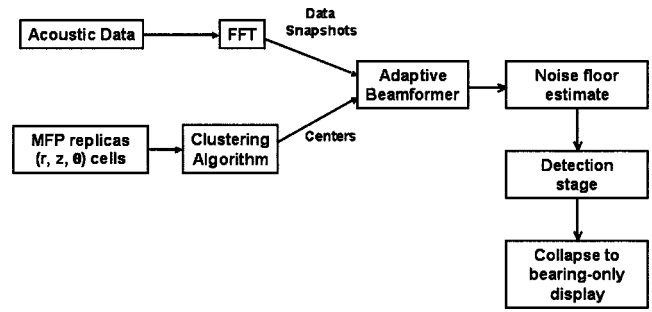
FIG. 4. Proposed signal processing flow for sampled field processing.

ing bearing only.

IV. USE OF CLUSTERS IN PROCESSING

Figure 4 shows a proposed signal processing flow that is motivated by the two previous sections. Those sections introduced the ideas of using clustering (to gain a nonredundant representation of the set of MFP replicas), and of bearings-only localization (to reduce sensitivity to environmental uncertainty and low SNR). The processing chain in Fig. 4 demonstrates how these two ideas can be combined. The steps in processing are as follows.

A. Beamforming stages

The five left-most boxes in Fig. 4 show steps related to beamforming. The steps involved in calculating and clustering the MFP replicas were described in Sec. II. Since the replicas depend on the environment and the array geometry, they need to be recalculated whenever either of these change. Clusters are generated for all frequencies of interest.

The acoustic data collected on the N array elements are transformed into the frequency domain using a short-time Fourier transform. This gives a set of $N \times 1$ data snapshots versus time and frequency. The snapshots are denoted $\mathbf{x}(f_0, t_0)$, where f_0 is the frequency and t_0 is the time of the snapshot. The data shown below were processed using a 2 s Hanning-windowed fast Fourier transform (FFT) with 50% overlap.

The nonadaptive or *conventional* beamformer (CBF) uses a weight vector that is simply a scaled version of the steering vector $\mathbf{w}_c(\cdot) = \mathbf{v}(\cdot)/N$. This normalization ensures that there is unity gain in the look direction. The conventional beamformer output for the direction Θ is given by

$$P_c(f_0, t_0, \Theta) = |\mathbf{w}_c^H(f_0, \Theta)\mathbf{x}(f_0, t_0)|^2. \qquad (6)$$

For simplicity, the dependence on the frequency f_0 and snapshot time t_0 is omitted below.

Adaptive beamforming (ABF) is implemented for the data analysis below using the dominant mode rejection (DMR) algorithm,[19] with a white noise gain constraint added to increase robustness to signal mismatch. DMR is a variant of the minimum variance distortionless response (MVDR) algorithm with diagonal loading. The loaded MVDR weight vector is given by

$$\mathbf{w}_d(\Theta) = \frac{(\hat{\mathbf{R}} + \delta(\Theta)\mathbf{I})^{-1}\mathbf{v}(\Theta)}{\mathbf{v}^H(\Theta)(\hat{\mathbf{R}} + \delta(\Theta)\mathbf{I})^{-1}\mathbf{v}(\Theta)}, \tag{7}$$

where Θ is the beam steering direction, $\delta(\Theta)$ is the direction-dependent diagonal load level, and $\hat{\mathbf{R}}$ is the estimated data covariance matrix. $\hat{\mathbf{R}}$ can be estimated from the data using a time-averaged sample covariance matrix of frequency-domain snapshots

$$\hat{\mathbf{R}} = \frac{1}{L}\mathbf{X}\mathbf{X}^H, \tag{8}$$

where $\mathbf{X}$ is a $N \times L$ data matrix, formed by accumulating L frequency-domain snapshots derived from short-time FFT data on the N sensors. Full-rank processing would require averaging over a long time period $(L \geq N)$, during which motion effects may become significant. Therefore, DMR focuses on the D largest eigenvalues (and corresponding eigenvectors) and deemphasizes the remaining $N\text{-}D$ eigenvalues/eigenvectors by constructing a modified sample covariance matrix $\tilde{\mathbf{R}}$

$$\tilde{\mathbf{R}} = \sum_{k=1}^{D} \lambda_k \mathbf{u}_k \mathbf{u}_k^H + \alpha \sum_{k=D+1}^{N} \mathbf{u}_k \mathbf{u}_k^H, \tag{9}$$

where α is an estimate of the input noise floor. $\tilde{\mathbf{R}}$ is then substituted for $\hat{\mathbf{R}}$ in Eq. (7) to calculate the adaptive weights $\mathbf{w}_d$. These weights are used in Eq. (6) (in place of $\mathbf{w}_c$) to find the adaptive beamformer output power P_a.

The load level δ in Eq. (7) is increased via an iterative process until it satisfies a white noise gain constraint[20]

$$\mathbf{w}_m^H(\cdot)\mathbf{w}_m(\cdot) \leq \frac{\beta}{N}. \tag{10}$$

The constant β above is a "relaxation" parameter, defined as the factor by which the weight norm in Eq. (10) is allowed to to exceed the "white noise gain" of $1/N$ (which is the value of the weight norm without diagonal loading and with white noise as input).

In passive processing applications, such as considered here, data from the signal of interest cannot be excluded when estimating the covariance matrix $\tilde{\mathbf{R}}$. An adaptive beamformer will (mistakenly) interpret any mismatched[21] portions of the signal of interest as noise and will (mistakenly) adapt to and reject this portion of the signal itself, an effect termed *signal self-cancellation* or *signal self-nulling*.[22] Diagonal loading via the white noise gain constraint provides a means for limiting this signal self-cancellation.[22] The white noise gain parameter β in Eq. (10) was set to 2 (3 dB) for the results below, yielding an adaptive processor that is relatively tolerant to signal model mismatch.

B. Noise estimation

After beamforming, the output powers are analyzed to estimate the noise level as a function of bearing. The harmonic mean noise estimator is used here.[23] The noise as a function of bearing θ is estimated as

$$P_N(\theta) = \left(\sum_{k \in \mathcal{K}} P(k)^{-1} \right)^{-1} \tag{11}$$

where $P(k)$ is the beamformer output power (either conventional or adaptive) for the kth cluster, and $\mathcal{K}$ is the set of clusters whose cluster center lies along a bearing within $\pm M$ azimuthal beams of θ. For data processing shown below, $M = 7$. The harmonic mean acts to suppress high-power outputs due to signals so that an accurate noise estimate is obtained. A variety of noise estimation methods are available, and the choice of the harmonic mean estimator is not critical for cluster-based processing.

C. Collapsing to bearing

Two possible methods for collapsing the beamformer outputs to bearing are considered here.

1. Maximize over all outputs along each bearing (ORing)

The simplest way to collapse MFP outputs to bearing only is to maximize over all matched field outputs along each bearing:

$$P_{1-D,mfp}(\theta) = \max_{(r,z)} P(r,z,\theta). \tag{12}$$

A similar collapsing can be done for a cluster-based beamformer. In this case the maximization is over the set $\mathcal{I}$ of clusters whose centers lie on the given bearing

$$P_{\text{SFP-OR}}(\theta) = \max_{i \in \mathcal{I}} P_i(\theta). \tag{13}$$

Because clustering captures the multipath structures described by the MFP replicas, the outputs of Eqs. (12) and (13) are expected to be very similar.

Signal power is preserved when using maximization to collapse beamformer output to bearing only. Unfortunately, maximization has the effect of increasing the noise level. As a thought experiment, imagine that all outputs along a given bearing correspond to noise. Maximizing the outputs gives a result that corresponds to the highest observed noise value, which may be significantly higher than the average noise level. As a result, a signal on an adjacent bearing will have a lower observed signal-to-background ratio. Maximization also tends to emphasize sidelobe energy from nearby sources, which again reduces the signal-to-background ratio.

The operation shown in Eqs. (12) and (13) has been labeled "ORing" in the sonar literature, and the SNR losses it causes are known as "ORing loss."[24,25] This terminology is adopted below. Data results generated using Eq. (12) will be labeled "ORed MFP," while cluster-based results generated using Eq. (13) will be denoted "SFP-OR."

2. Cluster-based detection approach

Clustering provides additional information about the relationship among points in the search space that is not available in standard MFP. The distance in terms of mismatch between cluster centers can be calculated, giving a measure

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Tracey, Lee, and Turaga: Sampled field processing 2639
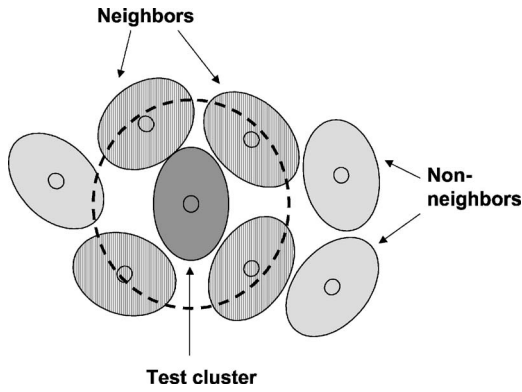
FIG. 5. Cartoon showing the proposed peak detection stage. Each cluster is compared to its "neighbors," defined as other clusters that are highly correlated with it. If the test cluster is a local peak or close to the local peak, its output is retained; if not, the cluster power is set to the estimated noise floor. The full set of clusters is then collapsed to bearings only.

of how different regions in the search space are related. A relatively simple method for using this information is described here.

One way to separate signals from noise is to require that a signal should give the peak output power in its local neighborhood. If a test cluster has output power P_{tst}, this cluster may be considered signal-like if

$$P_{\text{tst}} > P_j \quad \forall\, j \in \mathcal{N}_{\text{NB}}, \tag{14}$$

where j is the set of cluster centers in the neighborhood $\mathcal{N}_{\text{NB}}$. While noise outputs will occasionally give local peaks in output power, the majority of noise outputs will not. Noise-generated peaks will be spaced randomly in bearing and will not generate apparent source tracks over time. Sidelobe energy will also be suppressed using this test.

An approach to defining the local neighborhood $\mathcal{N}_{\text{NB}}$ is shown in Fig. 5. The neighborhood is defined as the set of all clusters whose centers are mismatched from the test cluster's center by less than some threshold

$$\text{dist}(\mathbf{v}_{\text{cen}}^{\text{tst}}, \mathbf{v}_{\text{cen}}^{j}) < = \eta_{\text{NB}}, \tag{15}$$

where distance is as defined in Eq. (5). Good results on data have been obtained by setting η_{NB} in the range of 2–3 dB, for cluster radii ranging between 0.5 and 1.5 dB. Typically the cluster neighbors identified as above extend over several adjacent bearings, and include some but not all of the clusters for each bearing.

In practice, a modification to this approach has been found to be helpful. Neighboring clusters that are not locally maximal [as defined in Eq. (14)] but have output power within 1 dB of the local maximum are retained. This modification helps in the situation where a source has energy that is split between two clusters.

The results of this detection stage are collected in a vector δ, whose entries are "1" for clusters that are found to be signal-like and "0" for clusters that are not. The full set of clusters is collapsed to bearing as

$$P_{\text{SFP-DET}}(\theta) = \max_{i \in \mathcal{I}} [P_i(\theta) \cdot \delta_i] + P_N(\theta)[1 - \max_{i \in \mathcal{I}} (\delta_i)], \tag{16}$$

where $\mathcal{I}$ is the set of clustered outputs whose cluster centers lie along θ. Thus, if there are one or more signal-like clusters along the bearing θ, the power output is set to the largest of them. If no signal-like clusters are identified, the power is set to the estimated noise floor.

A summary of the proposed processing method is as follows. Note that the first two steps below are also part of generating the ORed output described in Eq. (13) (SFP-OR).

- The full set of MFP replicas (steering vectors) is clustered, and the cluster centers are used as sampled field replicas.
- The replicas are used in beamforming. Cluster output powers and replica vectors are stored.
- An estimate of the noise floor versus bearing is made using the full set of cluster outputs.
- Each cluster's output power is examined in turn. If this power is higher than the power on any nearby cluster, the cluster is declared to be signal-like. Any neighbors with output powers that are within 1 dB of the peak are also labeled as signal-like.
- A maximization is used to collapse all signal-like clusters on each bearing to a bearings-only display. If no clusters along an azimuth have been labeled as signal-like, the output is set equal to the estimated noise floor.

Data results generated using this approach will be labeled "SFP-DET."

V. RESULTS USING EXPERIMENTAL DATA

Planewave, MFP, and sampled field approaches were tested using data from the SWellEx-96 ocean acoustics experiment (www.mpl.ucsd.edu/swellex96). Data were processed for event S5, in which the towship R/V Sproul towed both deep and shallow sources. These sources each projected a sequence of tones of varying levels. Results are shown for the bottom-mounted HLA North array, an irregularly spaced array of roughly 240 m overall length (see Fig. 6).

The sparse sampling provided by HLA North resulted in increased azimuthal sidelobes. Fig. 6(b) shows the calculated planewave response of HLA North to an endfire source at 148 Hz, one of the tones transmitted by the Sproul. As compared to a fully populated array, HLA North displays significantly elevated sidelobe levels.

The data from event S5 were beamformed using single-path (planewave and range-focused) replicas, MFP replicas, and cluster centers. Identical beamforming parameters were used for all replica types. The data were processed using 2 s FFT windows with 50% overlap. No spatial taper was applied to the array. Beams were linearly spaced in wave number, with 89 beams ranging from −70° to 70°.

Adaptive beamforming (ABF) was implemented using the DMR algorithm described above. As noted, the white noise gain parameter was set to 3 dB to give tolerance against mismatch. In the beamforming, 15 FFT snapshots were averaged in time to estimate the eigenvectors and ei-
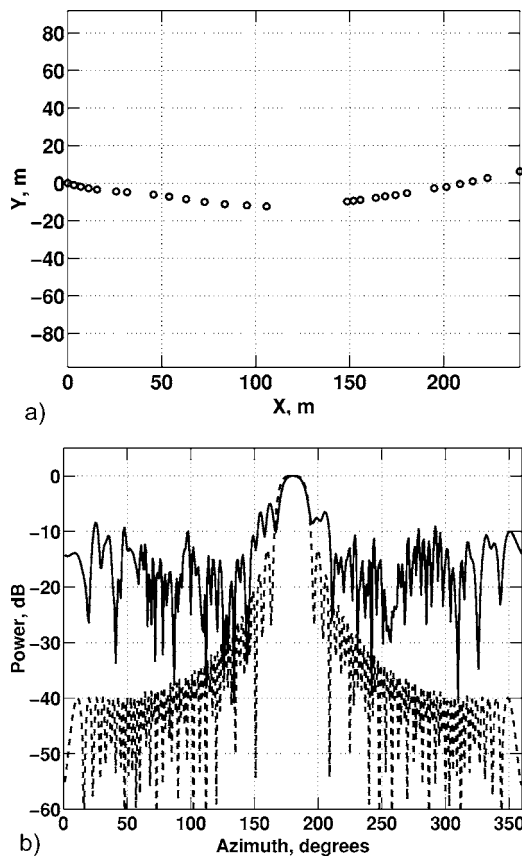
FIG. 6. Array geometry and beam patterns for bottom-mounted HLA North array from SWellEx-96. The sparse sampling causes elevates sidelobes (solid line) as compared to those expected for a fully populated array of the same length (dashed line). (a) Hydrophone locations and (b) beam patterns at 148 GHz.

FIG. 7. SWellEx-96 bearing-time records (BTRs), 235 Hz projected tone, for event S5. Cosine-spaced planewave beams are used for beamforming. The R/V Sproul was at endfire to the array during the times shown and transmitted tones continuously throughout the event. Note that adaptive beamforming suppresses sidelobe energy at the cost of signal self-nulling. (a) Conventional planewave and (b) adaptive planewave.

genvalues corresponding to the $D=5$ strongest sources in the data. The noise floor for the adaptive beamformer was found by averaging the remaining eigenvalues ($D+1$ through N, where N is the number of array elements).

A. Example planewave and ORed full-field results

This section gives example results for planewave and ORed full-field (MFP and clustered) beamformers. The examples help to illustrate the differences between the beamformers and serve as a preface to the more detailed data analysis in Secs. V B–V D.

Figure 7 shows planewave output versus bearing and time for the 234.5 Hz frequency bin, which contains most of the energy for the 235 Hz tone projected by the deep source. The towed source is at endfire, or 0°. The planewave conventional beamforming result, seen in Fig. 7(a), shows a large amount of energy in grating lobes caused by the array sparsity. Planewave adaptive beamforming, shown in Fig. 7(b), largely rejects this grating lobe energy.

The adaptive beamformer also gives much lower energy on the target track. Since both conventional and adaptive beamformers are normalized to give unit gain on target, one would expect equal power outputs if no mismatch were present. The reduction in along-track target power therefore

represents adaptive self-nulling caused by mismatch between the received signal and the assumed (planewave) signal model.

Figure 8 shows results of applying ORed full-field replicas to the same data. Figure 8(a) shows ORed MFP output [collapsed to bearing using Eq. (12)] while Fig. 8(b) shows SFP-OR output [collapsed to bearing using Eq. (13)]. A cluster radius of 1 dB was used for SFP-OR results. The ORed MFP and clustered beamformer results are very similar. The clustered output shows slightly more target self-nulling due to the fact that clustering uses a coarser sampling of the information contained in the MFP replica field.

A comparison of the towed source track (at endfire, or 0°) in Figs. 7 and 8 shows that use of full-field replicas gives significantly reduced adaptive self-nulling as compared to direct-path beamforming. Figure 8 also shows an increase in the displayed noise floor, or background, as compared with planewave output. This ORing loss reduces the overall signal-to-background ratio and motivates the use of alternate collapsing strategies.

In Figs. 7 and 8 a second apparent track is seen, initially to the right of the main track but moving quickly to the left. The source of this energy is unclear, and the source is not seen in other frequency bins. The main point of interest in Figs. 7 and 8 is the reduced adaptive self-nulling seen on the actual endfire target when full-field replicas are used.
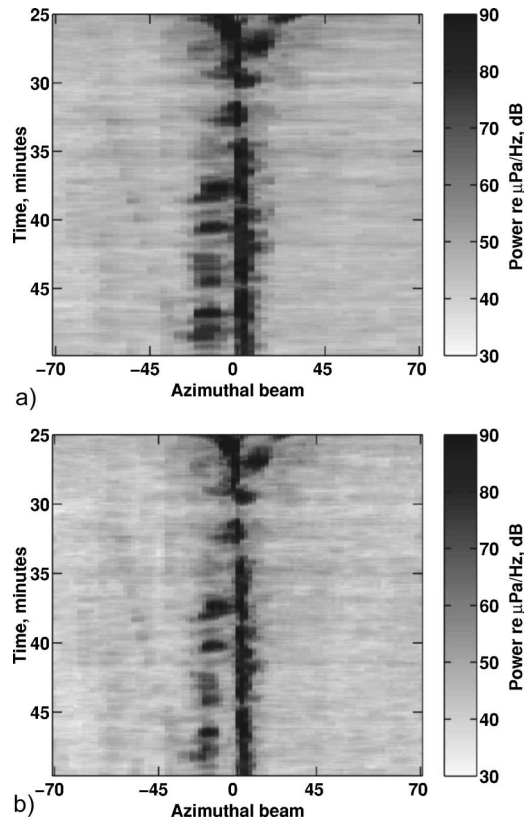
FIG. 8. ORed adaptive MFP and SFP-OR results for the 235 Hz projected tone, collapsed to bearing using Eqs. (12) and (13). Note the increases in both signal and noise levels relative to adaptive planewave results. Little difference exists between sampled field and MFP outputs when this collapsing strategy is used. (a) Adaptive MFP, ORed to bearing and (b) adaptive SFP-OR.

B. Measured mismatch reduction

This section quantifies the reduction in mismatch obtainable using clustered full-field replicas. Mismatch reduction was measured experimentally by using high-SNR tones broadcast by the R/V Sproul as calibration signals. By comparing the maximum beamformer output to the power of the signal received on the array, beamformer mismatch or signal gain degradation (SGD) can be measured as

$$\text{SGD} = \max_i \frac{\|\mathbf{v}_i^H \cdot \mathbf{x}\|^2}{\|\mathbf{v}_i\|^2 \|\mathbf{x}\|^2}, \tag{17}$$

where $\mathbf{v}_i$ are the set of replica vectors and $\mathbf{x}$ is the data, assumed to come from a single source. To verify that a single source dominated the data (and thus the mismatch could be accurately computed), a singular value decomposition was calculated versus time for a covariance matrix estimated by averaging four snapshots. The ratio of the first to second eigenvalues was computed, and SGD estimates were judged to be valid if this ratio exceeded 6 dB.

Figure 9 shows computed SGD versus time for the 234.5 Hz bin. SGD values are shown for planewave replicas and for clustered beamformer outputs computed using different cluster radii. A noticeable mismatch reduction is seen when moving from planewave to clustered replicas. This mismatch reduction is equal to the gain in output power expected for conventional beamforming.
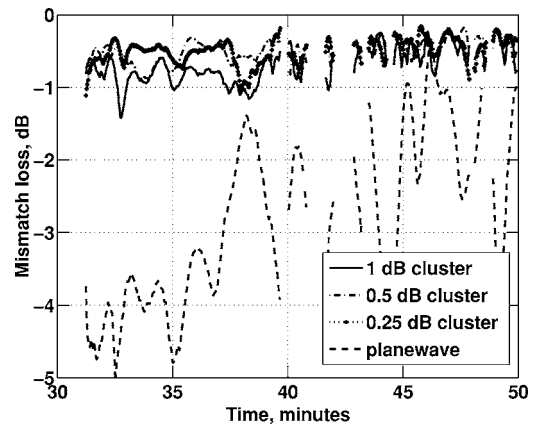
FIG. 9. Measured signal gain degradation vs time for 231.5 Hz frequency bin. Results are shown for planewave replicas and clustered full-field replicas. Results for cluster sizes of 1, 0.5, and 0.25 dB are shown. The clustered full-field replicas provide significant mismatch reductions as compared to planewave replicas.

Some mismatch reduction is observed when changing from a 1 dB cluster radius to a 0.5 dB cluster radius, while little additional reduction is seen when changing from 0.5 to 0.25 dB radii. This indicates that the inputs to the beamformer do not support a more detailed sampling of the pressure field. The remaining uncertainty is likely a combination of errors in estimated array positions and errors in the environmental modeling. This is an example of how detection gains are limited by environmental uncertainty.

Signal gain degradation values were calculated for four strong tones. Normalized histograms of the valid SGD values were generated for planewave and clustered replicas, using a 0.5 dB cluster radius. The resulting estimated probability density functions are shown in Fig. 10. The mismatch grows with frequency, as expected, but is noticeably lower for the clustered replicas. For the lowest three tones, the clustered replicas nearly always have 1.5 dB or less mismatch, while the planewave replicas can experience 2–4 dB mismatch. This dB difference can be greatly amplified in adaptive processing. Reducing mismatch by even a few decibels can dra-
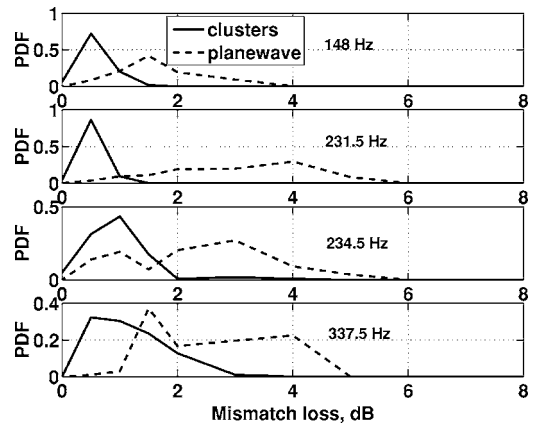
FIG. 10. Probability density of measured signal gain degradation as a function of frequency. Mismatch losses using MFP-derived clusters are always less than 3 dB, while planewave mismatch can be considerably higher.
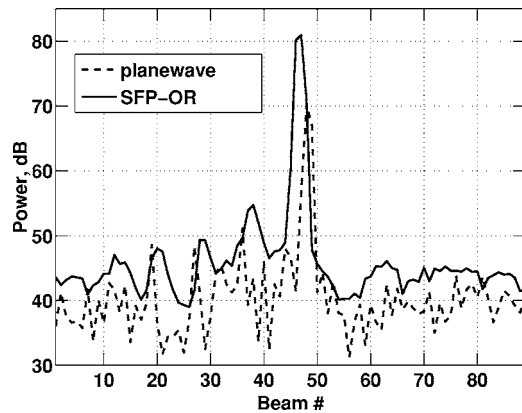
FIG. 11. Beamformer outputs at a single snapshot, comparing planewave adaptive beamfomer output to adaptive sampled field processing with simple ORing used to collapse to bearings only (SFP-OR). Mismatch reduction leads to less self-nulling and more signal power in the SFP-OR output, but ORing increases the displayed background level.

matically reduce adaptive self-nulling for strong sources, with smaller but still significant effects seen for weaker sources.

C. Comparison of collapsing options in data

As discussed above, collapsing MFP or clustered outputs to bearing using a simple maximization (ORing) leads to an increase in the apparent noise floor. Figure 11 illustrates this problem in more detail. Outputs of the planewave and ORed sampled field beamformers are shown versus bearing at a single snapshot in time. A cluster radius of 1 dB was used for this and all subsequent SFP results. The mismatch reduction obtained by SFP roughly 15 dB signal gain in this snapshot due to reduced target self-nulling (note that a weaker source would suffer less self-nulling loss). However, a noticeable increase in the quiet beams (away from the source) is also seen. This increase is referred to as *ORing loss*.[24]

Figure 12 compares noise floors estimated from planewave outputs, from the ORed sampled field outputs (SFP-OR), and from the *full* set of cluster outputs (SFP-DET noise floor, found using Eq. 11). This last noise estimate is used in Eq. (16) to generate the SFP-DET outputs. Visually, the SFP-
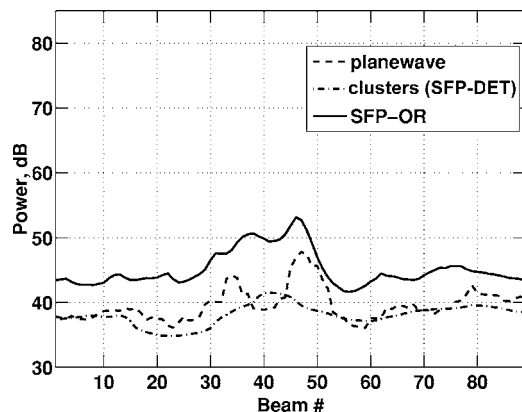
FIG. 12. Estimated noise floors for planewave adaptive beamforming, all clusters (SFP-DET noise floor), and SFP-OR results. The SFP-DET noise floor is similar to that of planewave beamforming, but ORing leads to a 5.5 dB increase in the average noise floor of SFP-OR.
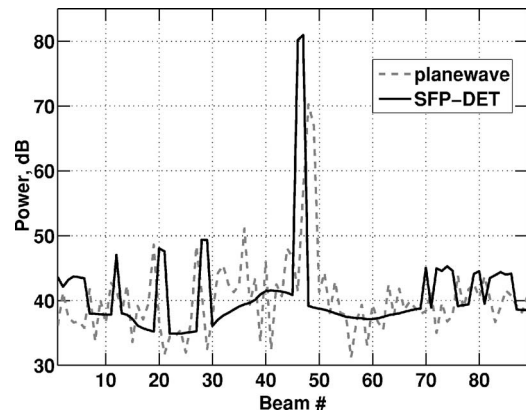
FIG. 13. Sampled field outputs using cluster-based detection (SFP-DET) compared to planewave ABF. For bearings that do not contain an identified peak in sampled field outputs, the SFP-DET output is set to the estimated noise floor for all clusters. The resulting display maintains signal gains due to mismatch reduction but minimizes the increase in the noise floor.

DET noise floor estimated from the full set of clusters appears close to the planewave noise level. The average noise floor for the planewave curve is 39.7 dB, while the average noise floor for SFP-DET is 38.1 dB; the small difference is probably due to spillover of signal energy into the planewave noise floor estimate. In contrast, the average noise estimate for the SFP-OR outputs is 45.2 or 5.5 dB higher than the planewave output. This increase is due to the ORing operation of Eq. (13). Figure 12 suggests that the SFP-DET algorithm has the potential to produce a similar noise floor to that of planewave beamforming.

Figure 13 shows the result of applying the SFP-DET approach. The planewave ABF result is shown as a dotted line, while the solid line shows the SFP-DET cluster result. For bearings where no cluster peak was found, the SFP-DET curve shows the underlying noise floor estimate plotted in Fig. 12. The SFP-DET result retains gains on the signal (which is at 0°) while avoiding a large increase in the noise floor. A number of small peaks are seen in the SFP-DET output, which may correspond to either local noise peaks or weak signals. The local SNR for these detections is of the same order of magnitude as similar noiselike peaks seen in the planewave output. An important note is that peaks generated by noise fluctuations will vary randomly, and are not expected to generate consistent tracks over time.

D. Single-tone bearing-time records

Planewave and sampled field beamformers can be usefully compared by forming bearing versus time output plots (BTR's). Results are presented for here both high-SNR and lower-SNR projected tones. Adaptive beamforming (ABF) is used in all cases. Because adaptive self-nulling increases with source level, the gains from mismatch reduction when using SFP-DET are most striking in the loud tone data. However, significant improvements are also seen for the quieter tones.

Figure 14 shows single-frequency beamformer output for the 231 Hz shallow source projected tone. This tone and the 235 Hz result shown above were among the loudest tones broadcast, with source levels of 158 dB re μPa. Planewave
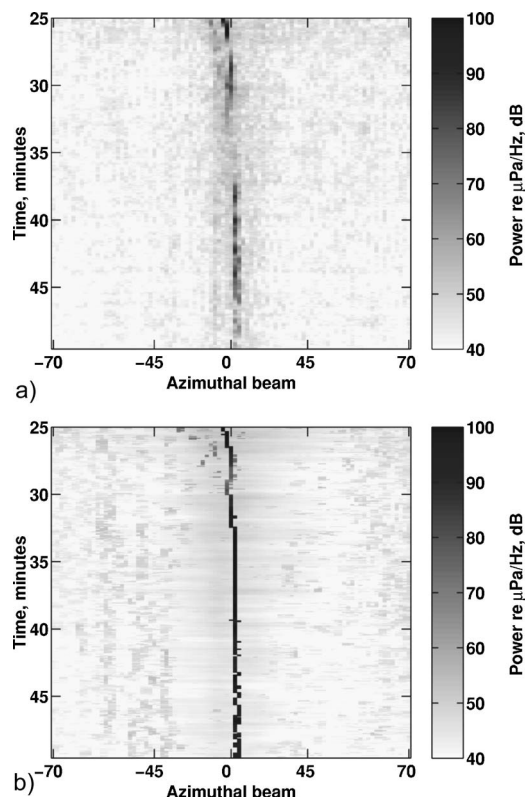
FIG. 14. SWellEx-96 bearing-time records (BTRs), 231 Hz projected tone, for time period with R/V Sproul near aft endfire (center of display). This tone was projected at 158 dB from the shallow source, which transited from roughly 2–4.5 km over the time period shown. Planewave beamforming is compared to SFP-DET sampled field results. Best results are seen with SFP-DET. (a) adaptive planewave and (b) adaptive SFP-DET.

results are shown in Fig. 14(a), while SFP-DET results [generated using Eq. (16)] are shown in Fig. 14(b). The high SNR towed source is visible near 0° in both displays, but the power output along the target track is much lower for the planewave beamformer than for the SFP-DET beamformer during several time periods. Since both beamformers are normalized to give unit gain on target, the difference in power output is attributable to adaptive self-nulling in the plane-wave result. The SFP-DET approach is successful in reducing the signal model mismatch so that the adaptive beamformer suppresses sidelobe energy without suppressing the signal of interest. In addition, the SFP-DET algorithm has a noise floor comparable to the planewave display, resulting in a clear improvement in signal-to-background ratio (SBR).

A similar result is seen in Fig. 15. This figure plots data from the 140 Hz FFT bin, which apparently contained a naturally radiated tone from the R/V Sproul (again near 0°). Note that these results are plotted on a 50 dB dynamic range, while plots above required larger dynamic range to accomodate the extremely high SNR. While the source level for this tone is unknown, it appears to lie in between the "loud" (158 dB) and "quiet" (132 dB) projected tones. Because this tone is not as strong as the 231 Hz projected tone, the reduction in self-nulling obtained by sampled field processing is not as extreme, though an improvement is clearly noticeable.

In Fig. 15, an additional subfigure (b) is added that shows a bearing-time record formed by maximizing over
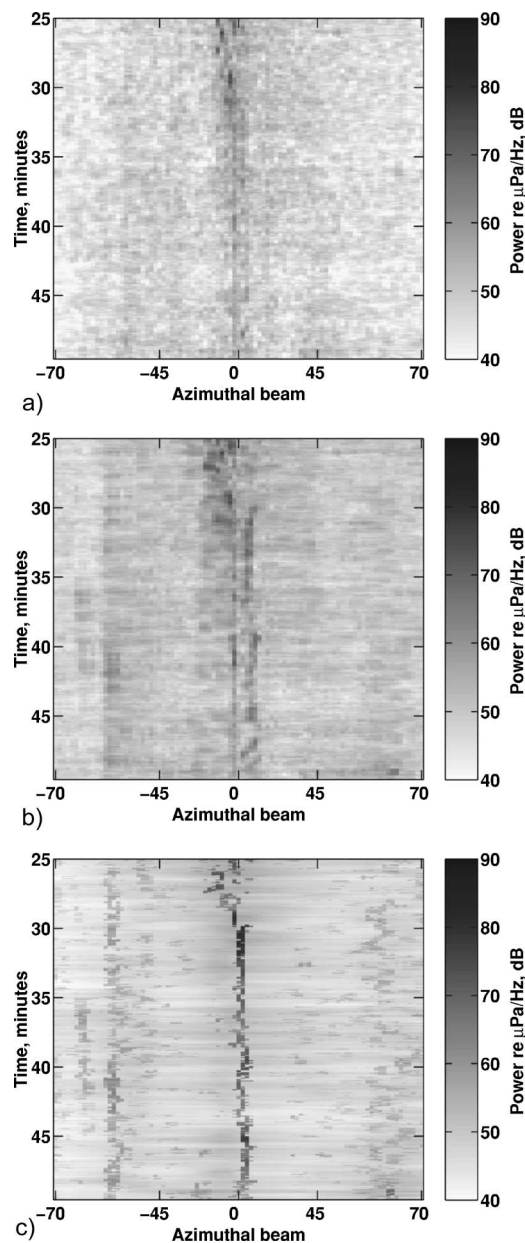
FIG. 15. Bearing-time records (BTRs) at 140 Hz. This tone appears to be an naturally radiated tone from the R/V Sproul. The SFP-DET sampled field result shows signal gains but maintains a low noise floor. The SFP-DET result performs better than either adaptive planewave or adaptive range-focused beamformers. (a) adaptive planewave, (b) adaptive range-focused, ORed to bearing, and (c) adaptive SFP-DET.

range-focused direct path beamformer output, using five focus ranges uniformly spaced between 2 and 10 km. This covers the same ranges as included in the full-field replicas. This plot was generated to demonstrate that the signal gains observed in SFP-DET outputs are due to the use of full-field replicas, rather the use of range focused replicas. Wavefront curvature is not present for endfire sources, where the towed source is located, so accounting for source range is not expected to affect gain on the towed source (although it does affect off-endfire sources).

Figure 16 shows results for a quieter (132 dB re μPa) projected tone. A number of off-endfire contacts are seen in the ORed range-focused and SFP results that are not seen in
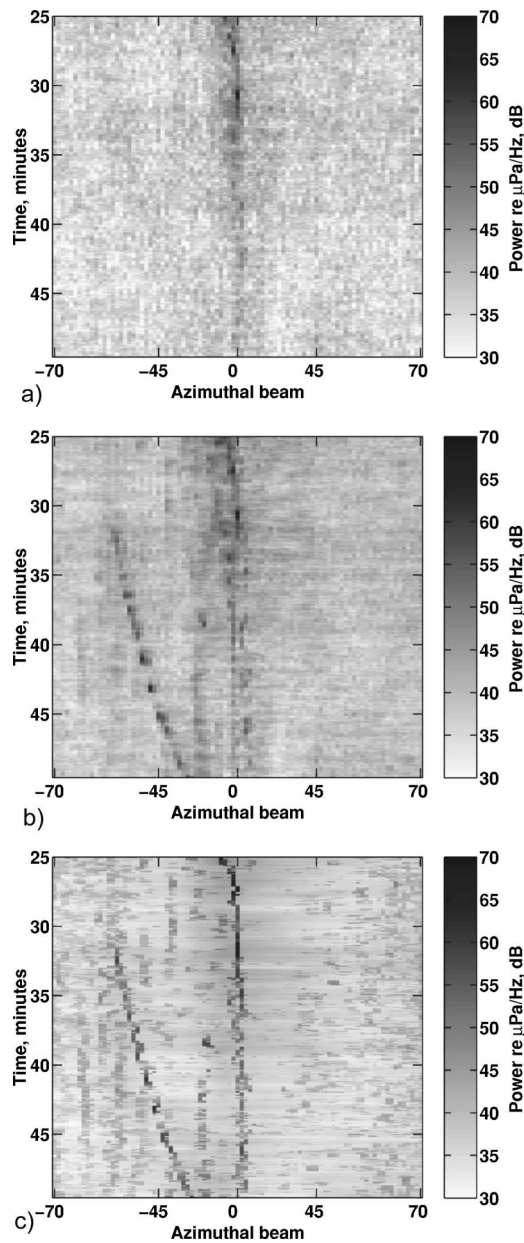
FIG. 16. Bearing-time records (BTRs) for the 237 Hz projected tone. This tone was projected from the source at 132 dB (one of the quieter projected tones). Several additional contacts are seen in the SFP-DET and range-focused outputs that are not seen in the planewave result, and are presumably due to closer-range sources. Again, the SFP-DET shows improved signal-to-background ratio on both the projected tones and observed contacts. (a) adaptive planewave, (b) adaptive range-focused, ORed to bearing, and (c) adaptive SFP-DET.

the planewave results. These are likely closer-range sources whose energy is protected when closer focus ranges are added. Results for the projected tone (near 0°) again show a noticeable signal-to-background gain for SFP-DET as compared to planewave or range-focused results.

This gain can be quantified in terms of the measured along-track signal-to-background ratio (SBR). Figure 17 shows estimated SBR for the ABF outputs from Fig. 16. The background level for each bearing of interest was estimated by averaging the output for three beams on either side, with an offset of two beams (thus the background for beam 30 is the mean of beams 25–27 and 33–35). SBR estimates were
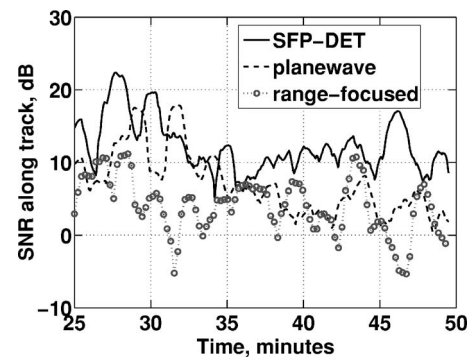
FIG. 17. Estimated along-track signal-to-background ratios (SBR), 237 Hz projected tone. Estimates are shown for adaptive SFP-DET, planewave, and ORed range-focused outputs. The median SFP-DET along-track SBR exceeds the median planewave SBR by 5.4 dB.

extracted along the observed signal track and smoothed using a 60 s moving average. The ORed range-focused result has the worst along-track SBR, presumably because the ORing operation causes an increase in background level. The SFP-DET approach has improved SBR as compared to planewave beamforming for almost all time intervals, giving an increase in median along-track signal-to-background ratio of roughly 5.4 dB. Higher gains would be expected if along-track SBR were calculated from Figs. 14 and 15, as the effects of adaptive self-nulling on planewave outputs are greater for these louder tones.

VI. SUMMARY

This paper has presented sampled field processing (SFP) as an alternative to matched field processing for exploiting the information about multipath structure provided by acoustic modeling. Unlike MFP, which attempts to simultaneously provide both detection gains and three-dimensional localization, SFP only seeks to provide detection gains. SFP uses cluster analysis to sample the predicted acoustic fields, resulting in a set of beamformer replicas that characterize typical multipath interference structures expected along each bearing. Because horizontal arrays can be used to estimate source bearing more robustly than range or depth, sensitivity to environmental uncertainty can be reduced by collapsing beamformer outputs to a bearings-only display. The information provided by cluster analysis is used in a detection scheme for collapsing beamformer outputs to bearing while preserving signal-to-background gains.

Gains resulting from application of this detection scheme (termed SFP-DET) were demonstrated using SWellEx-96 bottom-mounted horizontal line array data for an endfire source. Compared to adaptive planewave and adaptive range-focused beamformers, SFP-DET achieved better signal gain, because of its use of a full-field signal model. Compared to beamformers that collapse to bearings only by maximization (such as ORed MFP), the cluster-based SFP-DET approach achieved lower background levels. Overall, SFP-DET produced the best signal-to-background ratios of any of the beamformers considered.

The results shown here are for narrow-band (single-frequency) processing. Additional gains can be expected by

broadband processing, which combines signal energy across frequency. A simple form of broadband processing could be accomplished by incoherently averaging narrow-band BTRs such as those shown above. However, better performance would likely be achieved if the SFP-DET algorithm were modified to explicitly use broadband information. For example, a broadband version of Eq. (14) could be created which requires that a test cluster be labeled "signal-like" only if similar signal-like clusters are seen on the same bearing at other frequencies. Broadband processing using cluster analysis remains a topic for future work.

The work presented here has used cluster analysis as a preprocessing stage before beamforming is carried out. Alternatively, one could beamform using all MFP replicas, cluster the replicas corresponding to spatial peaks in the output to identify closely related peaks, and use a detection strategy to collapse these outputs to bearing. In effect, this would constitute a "beamform-cluster-detect" strategy rather than the "cluster-beamform-detect" strategy explored here. Thus, while the SFP approach examined here provided noticeable gains on experimental data, several avenues exist for further development.

ACKNOWLEDGMENTS

APPENDIX: EFFICIENT CLUSTERING ALGORITHMS

The computational problem for clustering replica vectors can be posed as follows: "Find the fewest number of cluster center vectors $\tilde{\mathbf{v}}^{\text{cen}}$ such that the maximum distance between each vector $\mathbf{v}_i$ and the closest center vector $\tilde{\mathbf{v}}_j^{\text{cen}}$ is less than η," where η is the cluster radius. This can be written

$$\max_i \min_j \text{dist}(\mathbf{v}_i, \tilde{\mathbf{v}}_j^{\text{cen}}) \leq \eta. \qquad (A1)$$

The exact solution to this problem is computationally prohibitive to calculate,[26] but approximate solutions can be found. The computational load of clustering full-field replicas was found to be greatly reduced for algorithms with the following characteristics: (1) the clustering method is "greedy," meaning that it will seek a suboptimal solution; and (2) a divide-and-conquer (denoted DnC) approach is used, in which the algorithm operates on subsets of replica vectors to build up the solution. For the cases studied, using a greedy algorithm rather than a more exact method (such as hierarchical clustering) gave a speedup of roughly $1000\times$. DnC gave an additional speedup of $10-20\times$.

A useful greedy clustering method has been proposed by Gonzalez.[26] This algorithm iteratively adds cluster centers by finding the vector that has the greatest distance from the previous set of centers. A simpler inner product clustering

algorithm was also developed. This algorithm first computes the $n \times n$ pairwise distance or mismatch matrix between every pair of vectors. The vector with the largest number of close neighbors, with mismatch less than the threshold, is picked as the cluster center. All replica vectors closer than the threshold are assigned to this center. The cluster center and cluster members are removed from the vectors to be clustered, and the process is repeated until all vectors are clustered. A final step re-assigns all non-centers to the closest center. The inner product method was found to be slightly faster than the Gonzales method and was used for the results shown above.

A divide-and-conquer (DnC) approach, in which the final clustering solution is constructed from the clustering solutions found for each subgroup of replicas, provides additional gains. DnC is implemented as follows: First, a subset of MFP replicas is extracted from the overall set. For the results shown above, every tenth replica was selected. A set of clusters is formed from this small set. A second subset of MFP replicas is extracted, and the projection of these new replicas onto the previous clusters is found. Any of the new replicas that fall within previously identified clusters are assigned to them; the remaining replicas are then clustered themselves, and the new clusters are added to the overall set of clusters. This process repeats until all MFP replicas are assigned to a cluster. Typically, the run time using this approach is dominated by the clustering of the first subset. Several different methods for breaking the MFP replicas up into subsets were examined. The subdivision method used did not have a strong effect on clustering performance.

[1] D. Manolakis, V. Ingle, and S. Kogon, *Statistical and Adaptive Signal Processing* (McGraw-Hill, New York, 2000).

[2] B. Tracey, "Statistical description of matched field processing ambiguity surfaces," J. Acoust. Soc. Am. **118**(3), 1372–1380 (2005).

[3] A. Baggeroer, W. Kuperman, and P. Mikhalevsky, "An overview of matched field methods of ocean acoustics," IEEE J. Ocean. Eng. **18**(4), 401–424 (1993).

[4] M. Siderius, M. Snellen, D. Simons, and R. Onken, "An environmental assessment in the Strait of Sicily: Measurement and analysis techniques for determining bottom and oceanographic properties," IEEE J. Ocean. Eng. **25**(3), 364–386 (2000).

[5] D. Knobles, R. Koch, L. Thompson, K. Focke, and P. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**(1), 205–222 (2003).

[6] M. Collins and W. Kuperman, "Focalization: Environmental focusing and source localization," J. Acoust. Soc. Am. **90**(3), 1410–1422 (1991).

[7] J. Krolik, "Matched-field minimum variance beamforming in a random ocean channel," J. Acoust. Soc. Am. **92**(3), 1408–1419 (1992).

[8] J. Tabrikian, J. Krolik, and H. Messer, "Robust maximum-likelihood source localization in an uncertain shallow-water waveguide," J. Acoust. Soc. Am. **101**(1), 241–249 (1997).

[9] H. Schmidt, A. Baggeroer, W. Kuperman, and E. Scheer, "Environmentally tolerant beamforming for high resolution matched field processing: Deterministic mismatch," J. Acoust. Soc. Am. **88**(4), 1851–1862 (1990).

[10] P. Daly, "Stochastic matched field processing for localization and nulling of acoustic sources," Ph.D. thesis, Massachusetts Institute of Technology, 2000.

[11] N. Lee and C. Richmond, "Threshold region performance prediction for adaptive matched field processing localization," *Proceedings, 12th Annual Adaptive Sensor Array Processing Workshop* (MIT Lincoln Laboratory, Lexington, MA, 2004) (www.ll.mit.edu/asap).

[12] B. S. Everitt, *Cluster Analysis* (Wiley, New York, 1993).

[13] B. Tracey, N. Lee, and S. Turaga, "Statistical clustering applied to adaptive matched field processing," *Proceedings, 12th Annual Adaptive Sensor*

Array Processing Workshop (MIT Lincoln Laboratory, Lexington, MA, 2004) (www.ll.mit.edu/asap).

[14] D. Rabideau, "Clutter and jammer multipath cancellation in airborne adaptive radar," IEEE Trans. Aerosp. Electron. Syst. **36**(2), 565–583 (2000).

[15] T. C. Yang and T. Yates, "Matched-beam processing: Application to a horizontal line array in shallow water," J. Acoust. Soc. Am. **104**(3), 1316–1330 (1998).

[16] F. Jensen, W. Kuperman, and M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1993).

[17] H. Cox, "Resolving power and sensitivity to mismatch of optimum processors," J. Acoust. Soc. Am. **54**(3), 771–785 (1973).

[18] L. Zurk, N. Lee, and J. Ward, "Source motion mitigation for adaptive matched field processing," J. Acoust. Soc. Am. **113**(5), 2719–2731 (2003).

[19] H. Cox and R. Pitre, "Robust DMR and multi-rate adaptive beamforming," *Proceedings of 31st Asilomar Conference* (1997), pp. 920–924.

[20] H. Cox, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **35**(1), 1365–1376 (1987).

[21] *Mismatch* is defined here as a difference between the assumed array response of a signal (i.e., its steering vector) and the actual array response in the data.

[22] N. Lee and N. Pulsone, "Performance of sample-covariance-based adaptive sonar detectors," *Proceedings of 34th Asilomar Conference* (2000), pp. 668-672.

[23] H. Cox and D. Pace, "A fast normalizer," *Proceedings of 30th Asilomar Conference* (1996), pp. 459–463.

[24] W. Struzinski, "ORing loss data for square law detectors followed by an ORing device and an accumulator," J. Acoust. Soc. Am. **105**(4), 2170–2180 (1999).

[25] G. Bottomley, "The effects of cross-correlated noise and multi-channel signal on ORing loss," *Proceedings, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. **2** (1987), pp. 26.6.1–26.6.4.

[26] T. F. Gonzalez, "Clustering to minimize the maximum inter-cluster distance," Theor. Comput. Sci. **38**, 293–306 (1985).

Continuous thermoacoustic mixture separation

G. W. Swift and D. A. Geller

Condensed Matter and Thermal Physics Group, Los Alamos National Laboratory, MS K764, Los Alamos, New Mexico 87545

The superposition of nonzero time-averaged mole flux $\dot{N}$ on a thermoacoustic wave in a binary gas mixture in a tube produces continuous mixture separation, in which one or more partially purified product streams are created from a feedstock stream. Significant product and feedstock flows occur through capillaries that are small enough to experience negligible thermoacoustic phenomena of their own. Experiments with a 50–50 helium-argon mixture show diverse consequences of nonzero flow, involving the addition of only one simple term $n_H \dot{N}$ to the equation for the heavy component's time-averaged mole flux, where n_H is the mole fraction of the heavy component. A boundary condition for n_H must be imposed on the equation wherever products flow out of the separation tube, but not where feedstock flows in. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2346129]

I. INTRODUCTION

In thermoacoustic mixture separation, a sound wave propagating in a gas mixture in a tube causes one component of the mixture to flow in the sound-propagation direction and the other component to flow in the opposite direction.[1–3] The expenditure of acoustic power results in an increase in the Gibbs free energy of the mixture's components, and the energy efficiency of the process is comparable to that of some other practical separation processes.[4] Thermoacoustic separation is caused by oscillating radial thermal diffusion combined with oscillating axial viscous motion. The sound wave's oscillating pressure causes an oscillating radial temperature gradient, which in turn causes radial thermal-diffusion oscillations, inducing a small fraction of the light and heavy components of the mixture to take turns being partially immobilized in the viscous boundary layer near the tube wall. Thus, outside of that boundary layer, the sound wave's axial oscillating motion carries gas that is slightly enriched in the heavy component in one direction during one half of the acoustic cycle and gas that is slightly enriched in the light component in the other direction during the other half of the cycle.[5] In a 2-m-long tube, a 50-50 helium-argon mixture has been separated to yield 70% helium, 30% argon at one end and 30% helium, 70% argon at the other end; and a small enrichment of ^{22}Ne from natural neon has been demonstrated.[6]

The five papers cited above described some of the mathematical physics of this phenomenon. Analytical expressions for the radial dependencies of oscillating temperature, density, mole fraction, and velocity and for the time-averaged molar separation flux were derived in the boundary-layer approximation, initially with no axial concentration gradient[1] and later with an arbitrary concentration gradient.[3] A numerical calculation was described that did not invoke the boundary-layer approximation.[3] Measurements confirmed the calculations, initially with no axial concentration gradient

in a tube where the boundary-layer approximation was adequate,[2] and later in a smaller-diameter tube with an axial concentration gradient.[3]

All of this theoretical work assumed that $\dot{N}_H + \dot{N}_L = 0$, where $\dot{N}_H$ and $\dot{N}_L$ are the time-averaged mole fluxes of the heavy and light components of the gas, respectively. In other words, it was assumed that the heavy and light mole fluxes were exactly equal in magnitude and in opposite directions. Similarly, all of the previous experimental work ensured that there was no bulk time-averaged total mole flux of the gas. In an engineering context, this corresponds to "batch" separation, in which a mixture is loaded into an apparatus, the apparatus is run for some time to accumulate an enriched product in one place and a depleted product in another place, and the products are finally removed. However, most industrial separation processes (e.g., petroleum refining, air separation, isotope enrichment[7–9]) run continuously, with feedstock steadily injected at one place while one or more enriched and depleted products are steadily removed elsewhere. Continuous separation requires nonzero total mole flux $\dot{N}$ superimposed on the separation flows, so that

$$\dot{N} = \dot{N}_H + \dot{N}_L \neq 0. \tag{1}$$

We undertook the work reported here to explore some fundamental aspects of continuous thermoacoustic mixture separation, for which inequality (1) is true.

Reference 3 showed that the first-order thermoacoustic momentum equation is unchanged by mixture-separation effects, and Doppler-shift effects are also negligible at the slow mean flows of interest here. Then the momentum equation shows that the evolution of the complex pressure amplitude p_1 with axial coordinate x is given by

$$\frac{dp_1}{dx} = -\frac{i\omega \rho_m / A}{1 - f_\nu} U_1, \tag{2}$$

where ω is the angular frequency of the oscillation, ρ_m is the mean density of the gas, A is the cross-sectional area of the

tube, U_1 is the complex volume-flow-rate amplitude, and f_ν is the viscous function[10] that depends on the geometry of the tube. The evolution of U_1 with x arises from the laterally averaged first-order continuity equation[10] for $dT_m/dx = 0$

$$i\omega\langle\rho_1\rangle + \rho_m\frac{d\langle u_1\rangle}{dx} = 0, \tag{3}$$

where $\langle\rangle$ indicates an average over the cross-sectional area of the tube. The ideal-gas equation of state gives

$$\frac{\langle\rho_1\rangle}{\rho_m} = \frac{p_1}{p_m} - \frac{\langle T_1\rangle}{T_m}, \tag{4}$$

and Eq. (30) in Ref. 3 with $f = \langle h\rangle$ gives $\langle T_1\rangle$. Combining these and using the ideal-gas identity $\rho_m a^2 = \gamma p_m$ yields the mixture-separation continuity equation

$$\frac{dU_1}{dx} = -\frac{i\omega A}{\rho_m a^2}\{1 + (\gamma-1)[Bf_\nu + Cf_{\kappa D} + (1 - B - C)f_{D\kappa}]\}p_1, \tag{5}$$

where a is the adiabatic sound speed, γ is the ratio of isobaric to isochoric specific heats, and B, C, $f_{\kappa D}$, and $f_{D\kappa}$ are defined in Ref. 3 and discussed in Appendix A here. Continuing through the derivation in Ref. 3, the first opportunity to incorporate nonzero $\dot{N}$ is in Eqs. (43) and (44) of Ref. 3, the latter becoming

$$\dot{N}_{H,m} = n_H\dot{N} - NAD_{12}\frac{dn_H}{dx} \tag{6}$$

with the addition of the convective term $n_H\dot{N}$, where n_H is the mean mole fraction of the heavy component and D_{12} is the binary mass-diffusion coefficient. Thus, the combination of Eq. (6) with Eqs. (41) and (56) of Ref. 3 shows that the mole flux of the heavy component is

$$\dot{N}_H = n_H\dot{N} + \frac{\delta_\kappa}{4r_h}\frac{\gamma-1}{\gamma}\frac{k_T}{R_{univ}T_m}|p_1||U_1|(F_{trav}\cos\theta$$
$$+ F_{stand}\sin\theta) + \frac{\delta_\kappa}{4r_h}\frac{\rho_m|U_1|^2}{m_{avg}\omega A}F_{grad}\frac{dn_H}{dx}$$
$$- \frac{\rho_m}{m_{avg}}AD_{12}\frac{dn_H}{dx}. \tag{7}$$

The first term represents the mole flux of the heavy component that simply accompanies nonzero bulk flow of the gas mixture. The second term represents the thermoacoustic mixture-separation process described qualitatively above, which can cause the heavy component to flow in either direction with respect to its concentration gradient. The third term represents a qualitatively similar process of oscillating radial mass diffusion and axial viscous motion, but one that always works to reduce the magnitude of a nonzero concentration gradient. The fourth term represents ordinary axial mass diffusion, which also always works to reduce the concentration gradient. In Eq. (7) and throughout this paper, $\delta_\kappa = \sqrt{2k/\omega\rho_m c_p}$ is the thermal penetration depth, k is the thermal conductivity, c_p is the isobaric specific heat, r_h is the hydraulic radius of the separation tube (for a circular tube, r_h is half of the circle's radius R), k_T is the molar

thermal-diffusion ratio, $R_{univ} = 8.314$ J/mol-K is the universal gas constant, T_m is the mean temperature, θ is the phase by which p_1 leads U_1, $m_{avg} = n_H m_H + (1 - n_H)m_L$ is the average molar mass, and m_H and m_L are the heavy and light molar masses, respectively. The three real variables F_{trav}, F_{stand}, and F_{grad}, which are generally <0, depend on the properties of the gas and the local geometry of the tube. Analytical expressions for the three F's have been published previously in the boundary-layer limit,[1,3] but in Appendix A here we present analytical expressions for the F's in small circular tubes for the first time. The coupled differential Eqs. (2), (5), and (7), with the expressions for the three F's given in Appendix A, have been implemented in the computer code DELTAE,[11] which has been used to create all the calculated curves in the figures below. In that implementation, Eq. (7) is solved for dn_H/dx and is integrated with respect to x simultaneously with Eqs. (2) and (5) to obtain $p_1(x)$, $U_1(x)$, and $n_H(x)$. The mole fluxes $\dot{N}$ and $\dot{N}_H$ are taken to be independent of x except where feedstock is injected or product is removed.

To solve Eqs. (2), (5), and (7) for $p_1(x)$, $U_1(x)$, and $n_H(x)$, boundary conditions on those three variables must be imposed. The boundary conditions for p_1 and U_1 are the same as for ordinary acoustics in tubes, but boundary conditions for n_H are less familiar. In batch separations, the number of moles of each gas in the apparatus at the end of the separation process is the same as at the beginning, so n_H must satisfy

$$\int n_H(x)\,dV = n_{H,batch}V, \tag{8}$$

where $n_{H,batch}$ is the initial heavy mole fraction of the batch and V is the volume of the apparatus. Continuous separations impose different constraints. In steady state, the number of moles of gas and the number of moles of the heavy component of the gas in the apparatus must be constant, so whatever flows in must flow out. This can be expressed mathematically as

$$\sum_j \dot{N}_{feedstock\,j} = \sum_k \dot{N}_{product\,k}, \tag{9}$$

$$\sum_j n_{H,feedstock\,j}\dot{N}_{feedstock\,j} = \sum_k n_{H,product\,k}\dot{N}_{product\,k}, \tag{10}$$

where the indices j and k account for more than one feedstock and product capillary, respectively. Equation (10) is a global constraint on n_H for continuous separation, similar in principle to the constraint of Eq. (8) for batch separation.

In addition to this global constraint, continuous separation imposes local boundary conditions. In a capillary connected to the separation tube at $x = x_o$ through which a product is removed continuously and which is small enough in diameter that acoustics within it is insignificant and the concentration gradient is zero, Eq. (7) is simply $\dot{N}_{H,product} = n_{H,product}\dot{N}_{product}$. The value of $n_{H,product}$ in that capillary must equal the local value in the separation tube where the capillary is connected. If the product is removed at an end of the separation tube, $\dot{N}$ and $\dot{N}_H$ in the capillary must also

equal their values in the separation tube. Thus, the boundary condition imposed on n_H at an end x_o of the separation tube where a product is removed is simply

$$\dot{N}_H(x_o) = n_H(x_o)\dot{N}(x_o). \tag{11}$$

(The corresponding condition would be

$$\dot{N}_H(x_o^-) - \dot{N}_H(x_o^+) = n_H[\dot{N}(x_o^-) - \dot{N}(x_o^+)] \tag{12}$$

if product were removed somewhere in the middle of the separation tube.)

The algebraic simplicity of the boundary condition given in Eq. (11) disguises at least two interpretations. At the simplest level, Eq. (11) expresses "what goes in must go out" with respect to the region of the separation tube adjacent to the product-removal capillary's entrance. However, Eqs. (7) and (11) together show that the last three terms in Eq. (7) must sum to zero at such a location. This shows that the boundary condition can also be interpreted as a complicated local constraint relating dn_H/dx (which appears explicitly in two of those three terms) to n_H itself (on which most of the variables in all three terms depend implicitly).

In contrast to continuous product removal, continuous feedstock injection does not impose any explicit local boundary condition on the solution to Eqs. (2), (5), and (7) in the separation tube. The total injection flux $\dot{N}_{feedstock}$ and the injection flux of the heavy component, $n_{H,feedstock}\dot{N}_{feedstock}$, add to any time-averaged flows already in the separation tube and compete with the terms of Eq. (7) and boundary conditions imposed elsewhere to determine the local value of n_H, which is not necessarily equal to $n_{H,feedstock}$. If the feedstock capillary had a large enough diameter to support internal thermoacoustics, continuity of n_H at the junction of the capillary and the separation tube would presumably hold. As its diameter was reduced, dn_H/dx at the end of the capillary would steepen, eventually developing what we conveniently treat as a discontinuity here.

The verification of the presence of the first term in Eq. (7), $n_H\dot{N}$, and the boundary condition on n_H expressed in Eq. (11) are the principal quantitative goals of this work. The experiments described below used a helium-argon mixture in a loudspeaker-driven traveling-wave separation tube. Fed with a 50-50 mixture, concentrations at the ends of the tube ranged from 30 to 70%. Feedstock was added and products were removed via capillaries, small enough in diameter to eliminate significant internal thermoacoustic effects but large enough to allow steady flow without requiring high pumping power. Feeding gas to the middle of the separation tube and extracting products at both ends allow simultaneous continuous production of helium-enriched gas and argon-enriched gas. Feeding gas to one end and removing an almost equal amount from the same end allows continuous production of a single more-highly enriched gas at the other end. Measurements under both circumstances provide quantitative confirmation of the $n_H\dot{N}$ term in Eq. (7) and of the boundary condition expressed by Eq. (11).
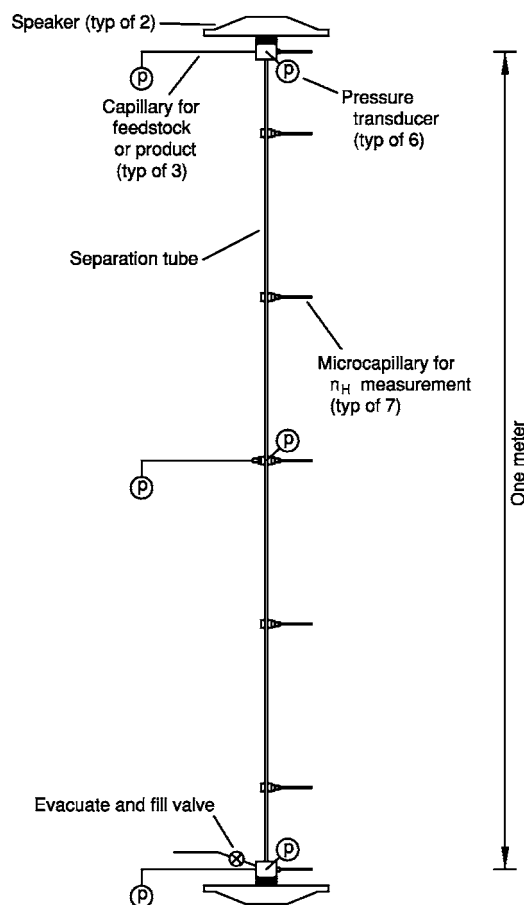
FIG. 1. Scale drawing of the apparatus. Valves to the right of the drawing, not shown, connected the microcapillaries to the RGA and pumps. Valves to the left of the drawing, not shown, connected the feedstock/product capillaries to the source of gas and other pumps.

II. APPARATUS AND INSTRUMENTATION

Figure 1 illustrates the apparatus we used to investigate these issues. The stainless-steel separation tube was 1-m long and had a 3.3-mm inside diameter. Near one end, access through a valve and a very short, 1-mm-inner-diameter (i.d.) connector allowed rapid evacuation and addition of the gas mixture before each experimental run, to a mean pressure $p_m = 80$ kPa that was measured with a high-accuracy capacitance manometer[12] that had been checked recently against a mercury manometer. At each end of the tube was a bellows-sealed piston epoxied to the face of a loudspeaker. These two driver assemblies were used to maintain the sound wave in the gas, and are described more thoroughly in Ref. 6. They were driven sinusoidally at 200 Hz by two phase-locked signal generators and audio amplifiers. Three piezoresistive pressure transducers[13] and a lock-in amplifier[14] were used to observe the resulting complex pressure amplitude p_1 in the gas at the ends and the middle of the separation tube. Table I summarizes the important dimensions of the apparatus discussed above and other variables that are discussed later.

The heavy mole fraction n_H in the separation tube could be measured at the seven locations that are shown in Fig. 1. At each location, a microcapillary[15] continuously withdrew gas from the separation tube at a rate of $\sim 3 \times 10^{-10}$ mol/s. All seven microcapillaries were 5-cm long and nominally

TABLE I. Properties of 50–50 helium-argon at 80 kPa, 300 K, and 200 Hz, and values for three tubes filled with that gas. The last three rows of the table are for $|p_{1,0}| = 2$ kPa.

	Microcapillary	Capillary	Separation tube
δ_ν	0.23 mm	← same	← same
δ_κ	0.36 mm	← same	← same
δ_D	0.39 mm	← same	← same
R	0.005 mm	0.064 mm	1.67 mm
$|\alpha|^{-1}$	3 mm	37 mm	
Δx	50 mm	150 mm	1000 mm
λ	2180 mm	← same	← same
F_{trav}	-9.43×10^{-12}	-5.48×10^{-6}	-0.345
F_{stand}	1.59×10^{-7}	4.58×10^{-4}	-0.047
F_{grad}	-7.11×10^{-8}	-2.05×10^{-4}	-0.210
Eq. (B10) at $x=0$	-6×10^{-7}	-0.02	
$|\Delta n_H|$	1×10^{-8}	3×10^{-4}	
$\dot{E}_2$	5 nW	10 μW	

10-μm inside diameter. These fragile microcapillaries were epoxied into larger copper tubes to protect against accidental damage. We found that two of them, henceforth referred to as the "large" microcapillaries, had 30% lower flow impedances than the other five, so the inside diameters of those two may have been 7% larger than the other five. The helium and argon flowing through the microcapillaries were detected by a computer-controlled, quadrupole-type residual gas analyzer[16] (RGA) pumped by a turbopump,[17] the latter backed by an oil-free mechanical pump.[18] A network of bellows valves allowed any of the seven microcapillaries to be connected to the RGA-pump system, which maintained the pressure at $\sim 10^{-6}$ torr in the presence of such flow. Meanwhile, the other six microcapillaries were connected to a second turbopump and second mechanical pump, to maintain high vacuum in all connecting tubing so that opening a valve never dumped a large amount of accumulated gas into the RGA. A measurement of n_H at a single microcapillary involved opening and closing the valves appropriately, waiting 2 min for steady state, and then averaging the RGA mass-4 and mass-40 "partial pressures" for 3 min. This averaging time yielded roughly 1% statistical uncertainty in the supposed ratio of mass-4 partial pressure to mass-40 partial pressure.

However, obtaining n_H from that "partial pressure" ratio was complicated because the RGA is much more sensitive to argon than to helium and because the RGA-pump system's response is nonlinear,[19] with the nonlinearity being different for mass 4 and mass 40. We mapped out this complexity near our operating conditions by recording the RGA's reported partial pressures for known helium-argon mixtures, which we prepared from the pure gases in a 2-liter fan-stirred mixing chamber. These mixtures, ranging in concentration from 35–65 to 65–35, were forced through one small and one large microcapillary with 80-kPa pressure. The RGA results were slightly different for the partial pressures resulting from the large and small microcapillaries. We found

$$\left(\frac{\text{pp}_4}{\text{pp}_{40}}\right)_{\text{true}} = C_L \left(\frac{\text{pp}_4}{\text{pp}_{40}}\right)_{\text{RGA}}, \quad \text{large microcapillaries;} \tag{13}$$

$$\left(\frac{\text{pp}_4}{\text{pp}_{40}}\right)_{\text{true}} = C_S \left(\frac{\text{pp}_4}{\text{pp}_{40}}\right)_{\text{RGA}}$$
$$\times \left[1 - 0.1\left(\frac{\text{pp}_4}{\text{pp}_{40}}\right)_{\text{RGA}}\right], \quad \text{small microcapillaries;} \tag{14}$$

where pp stands for the partial pressure reported by the RGA, $C_L \simeq 7.8$ and $C_S \simeq 8.0$. Based on a number of calibration runs, we suspect that systematic uncertainties in these values contribute about 1% to uncertainty in $(\text{pp}_4/\text{pp}_{40})_{\text{true}}$. The heavy mole fraction is then given by

$$n_H = \frac{1}{1 + (\text{pp}_4/\text{pp}_{40})_{\text{true}}}. \tag{15}$$

At the start of each experimental run (and at the end of the longest full-day runs), we remeasured C_L and C_S with our 50-50 feedstock. These constants changed from day to day by up to 2%. However, we retained the nonlinear coefficient 0.1 in Eq. (14) for all experimental runs.

In laminar flow of an ideal gas, the mole flux through a circular capillary of diameter D from pressure p to zero pressure is obtained from[20]

$$-\frac{dp}{dx} = \frac{128\mu}{\pi D^4} U, \tag{16}$$

where $U = \dot{N} R_{\text{univ}} T/p$ is the volume flow rate and μ is the viscosity. Integrating Eq. (16) from $x=0$ at one end of the capillary to $x=\Delta x$ at the other shows that

$$\frac{[p(0)]^2 - [p(\Delta x)]^2}{2} = \mu \frac{128\Delta x}{\pi D^4} \dot{N} R_{\text{univ}} T \equiv \mu Z \dot{N} R_{\text{univ}} T, \tag{17}$$

where $Z = 128\Delta x/\pi D^4$ includes the geometrical parts of the flow impedance. At the high-pressure end of a microcapillary, the mean free path is only about 1% of the diameter, so the gas should flow as a viscous fluid. However, $p(\Delta x) = 0$ at the low-pressure end, and the assumption of laminar flow is questionable because the mean free path exceeds the capillary diameter. The crossover to the molecular-flow regime, where the mean free path is of the

order of the microcapillary diameter, occurs around p = 0.01 bar, approximately 5 μm from the low-pressure ends of the 5-cm long microcapillaries. Safely ignoring such a small fraction of the microcapillary length, Eq. (17) shows that $\dot{N} \propto [p(0)]^2$. This quadratic dependence was experimentally observed, using the argon partial pressure at the RGA as a proxy proportional to $\dot{N}$. (This proportionality depends on the assumptions that the volumetric pumping speed of argon through the turbopump is independent of argon partial pressure, justified by the pump's specifications, and that the RGA's sensitivity to argon is approximately independent of partial pressure.)

Fittings at the top, middle, and bottom of the separation tube allowed connection of capillaries for feedstock injection and product removal. Figure 1 shows one such capillary at each of those three locations; other arrangements are discussed below as they arise. Each of these stainless-steel capillaries was nominally 127 μm (0.005 in.) inside diameter and 15-cm long. To obtain an accurate value of the geometrical factor Z for each capillary, we used an auxiliary setup in which we timed the displacement of a liquid out of an inverted graduated cylinder while feedstock gas with n_H = 0.50, driven by a known pressure, flowed steadily through the capillary into the graduated cylinder. A mixture of 90% glycerin, 10% water, and a few drops of food coloring provided low vapor pressure, not-too-high viscosity, and easy viewing. Care was taken to avoid effects of surface tension and to account for the changing head of the liquid. Equation (17) and the time derivative of the ideal-gas law, $pU = \dot{N}R_{univ}T$, were used to obtain Z from the measured pressures and volume flow rates. Five measurements were made with each capillary, with flow rates ranging from 1 to 5 μmol/s. No single determination of Z differed from the average for that capillary by 1%. We believe that systematic errors totaled at most 2%. (The measurements showed that the inside diameters of the capillaries were only 5 to 6% larger than their nominal values.)

Later, during mixture-separation experiments, the mole fluxes into and out of the separation tube through these capillaries were controlled by needle valves[21] connected to the feedstock gas through a pressure regulator[22] and to a vacuum pump through an old, homemade rubber-membrane manostat.[23] The flow rates were determined via Eq. (17) by measurements of the pressures at the ends of the capillaries using the piezoresistive transducers shown in Fig. 1 and knowledge of Z obtained as described in the previous paragraph. We accounted for the n_H and T dependences of the gas viscosity. We neglected acoustic streaming's contribution to the pressure difference[24] across the capillaries $|p_1|^2/4p_m$ because it was only 0.04 to 1% of the pressure difference of Eq. (17) for all measurements.

Establishing a desired wave in the separation tube was an iterative procedure, carried out intermittently over a period of an hour or two as steady state was approached. We adopted a standard target of 3.00 kPa amplitude in the center of the tube, and a purely real acoustic impedance $z = Ap_1/U_1$ equal to $\rho_m a$ in the center of the tube to achieve a pure traveling wave there. We made a DELTAE model of the
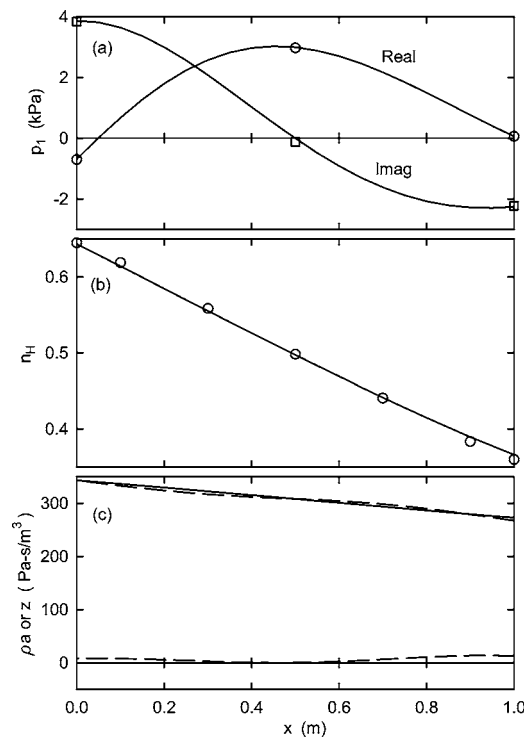
FIG. 2. Results with no time-averaged total mole flux, plotted as a function of location x along the separation tube. Curves are calculations and points are experimental values. (a) Complex pressure. (b) Heavy mole fraction. (c) Specific acoustic impedance z compared with $\rho_m a$. Upper dashed curve is Re[z] and lower dashed curve is Im[z]. Solid curve near Re[z] is $\rho_m a$.

apparatus with these constraints and with the desired conditions of steady flow. The DELTAE model's complex p_1 at the two ends of the separation tube then served as goals to which experimental measurements could be compared, and adjustments to the complex voltages imposed on the two-driver assemblies were made to bring the experimental p_1's close to their goals. This might seem to be a complicated procedure, because two experimental voltage amplitudes and two experimental voltage phases were being adjusted to try to meet two experimental pressure amplitudes and two experimental pressure phases. However, the overall time phase of the oscillation is a meaningless variable, and the overall amplitude of the wave was reliably changed by any desired fractional amount when the amplitudes of both drive voltages were changed by that same fractional amount. Thus, it was necessary only to measure changes in p_1 at the ends of the separation tube in response to two small changes: A small change in phase difference between the two driver voltages, and a small fractional increase in driver voltage at one driver. The matrix of these measurements was inverted to allow a prediction of the complex drive voltages needed to achieve desired p_1's. Two iterations might have sufficed to bring the wave close to the goal, but $n_H(x)$ was also evolving during the course of these iterations, changing the density and sound speed along the separation tube, so three or four iterations were usually needed.

To illustrate how well this wave-tune-up procedure was typically accomplished and to establish confidence in the measurement techniques, Fig. 2 shows experimental and DELTAE results with no steady flow in the separation tube.

Figure 2(a) compares the experimental and calculated pressure waves, with the zero of phase arbitrarily chosen to be the phase of the calculated p_1 at the middle of the separation tube. The agreement between the calculated and measured complex pressures at the two ends is indicative of how well the tune-up procedure described in the previous paragraph was routinely carried out, and the agreement between the calculated p_1 and measured p_1 at the middle of the separation tube is then one indication of how accurately we are modeling the wave. Figure 2(b) compares experimental and calculated mole fractions, and is a second indication of how accurately we are modeling the wave and its mixture-separation effects. The differences between measurements and calculations in Fig. 2(b) are about 1%, comparable to the expectations outlined above and similar to what was described in Ref. 3. Figure 2(c) shows calculations of specific acoustic impedance $z=p_1A/U_1$ along the tube for this wave, showing how well this arrangement can maintain a purely traveling wave along its entire length, even while the acoustic power decreases from 0.19 W at $x=0$ to 0.08 W at $x=1$ m.

Imposing nonzero mole flux on the separation tube through the feedstock and product capillaries requires a small amount of additional, nonacoustic power to overcome the viscous impedances of the capillaries. In the present experiment, the ideal mechanical power required to isothermally compress the feedstock from 80 kPa to the feedstock pressure and to isothermally compress the products from the product-suction pressure to 80 kPa can be taken as a fair measure of the additional power required. As an example from the high end of the range of flows discussed below, this power totals only 10 mW for 4 μmol/s flowing in through one feedstock capillary and 2 μmol/s flowing out through each of two product capillaries.

To accurately measure $n_H(x)$ with this apparatus, we must be assured that the concentration gradient in each microcapillary is small enough that n_H at the RGA does not differ significantly from n_H in the separation tube where the microcapillary is attached. To perform continuous separations, we must similarly be certain that dn_H/dx in the feedstock capillaries and product capillaries can be neglected, or at least is understood quantitatively. We must also know that acoustic power dissipation in the capillaries is acceptably small. In Appendix B, we analyze these issues for $\dot{N}=0$, using the small-radius limits of the mixture-separation results in Appendix A and other equations of thermoacoustics. The results are summarized in Table I. Without superimposed steady flow, the expected end-to-end mole-fraction differences $|\Delta n_H|$ would be approximately 10^{-8} in the microcapillaries and 3×10^{-4} in the feedstock/product capillaries, both much smaller than the mole-fraction resolution of our apparatus. Table I also shows that the acoustic power dissipated in these capillaries is negligible compared with the ~ 100 mW consumed in the separation tube.

However, in most of the circumstances of interest in this paper, the steady flow in the capillaries overwhelms the oscillating flow, invalidating the analysis in Appendix B. At 1 μmol/s in a feedstock/product capillary, the steady flow equals the amplitude of the oscillating flow at the high-$|p_1|$
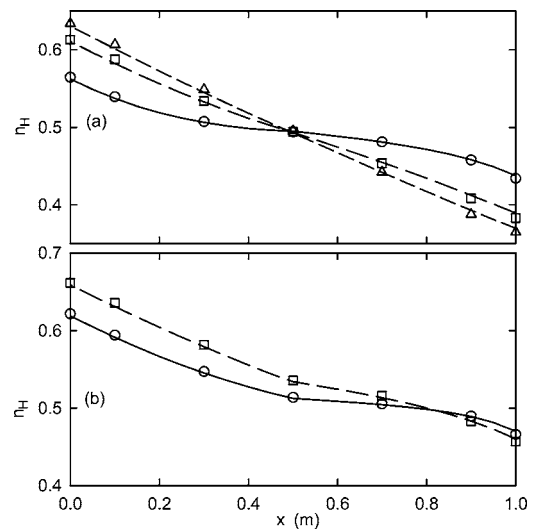
FIG. 3. Heavy mole fraction n_H as a function of location x, for feedstock injected at $x=0.5$ m and products removed at $x=0$ and $x=1$ m. Points are experimental values and associated curves are calculations. (a) Product mole fluxes equal. Circles, $\dot{N}_{\text{feedstock}}=3.96$ μmol/s. Squares, 0.99 μmol/s. Triangles, 0.26 μmol/s. (b) Product mole flux at $x=0$ is 25% of product mole flux at $x=1$ m. Circles, $\dot{N}_{\text{feedstock}}=3.63$ μmol/s. Squares, 1.59 μmol/s.

end of the capillary, and the capillary is swept clean by the steady flow in only ten acoustic cycles. In the microcapillaries, the steady flow is about twice the amplitude of the oscillating flow at the high-$|p_1|$ end. Thus, the analysis in Appendix B is probably only valid for flows below about 0.1 μmol/s in the feedstock/product capillaries. Nevertheless, it is hard to imagine how fast steady flows could *increase* the ability of the thermoacoustic phenomena to create a concentration gradient—in the next section, we see that modest steady flows always decrease the gradient. Two microcapillaries (not shown in Fig. 1) leading to the RGA from the left ends of the two product capillaries shown in Fig. 1 were used to verify that the product streams indeed carried the same mole fractions as were observed in the separation tube at the product-capillary entrances.

III. DISCUSSION OF THERMOACOUSTIC RESULTS

Figures 3–5 show the heavy mole fraction in the separation tube under a variety of conditions of nonzero total mole flux of feedstock and product(s). In all cases, calculations and measurements are in good agreement, confirming the presence of the $n_H\dot{N}$ term in Eq. (7) and our understanding of the boundary condition on n_H discussed near Eq. (11). It should be noted that there are no adjustable parameters in the calculations producing the curves in these figures. The calculations integrate Eq. (7), simultaneously with Eqs. (2) and (5), with respect to x, using the experimental values of $p_1(0)$, $p_1(x=1$ m), frequency, mean pressure, $n_{H,\text{feedstock}}$, and $\dot{N}$ in each feedstock and product capillary, and with the boundary condition Eq. (11) imposed on n_H wherever product flows out of the separation tube.

Figure 3 shows n_H versus x for feedstock injection at $x=0.5$ m. In Fig. 3(a), where half of the feedstock is removed at each end of the separation tube, the effect of simply in-
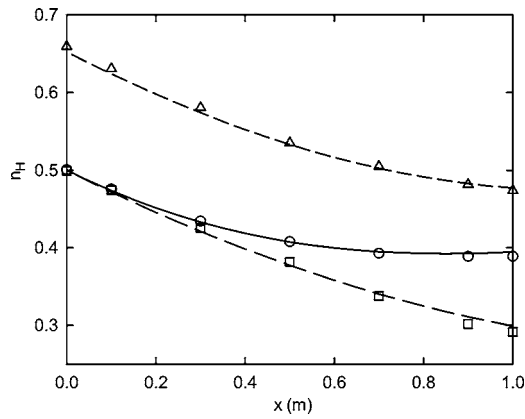
FIG. 4. Heavy mole fraction n_H as a function of location x, for feedstock injected at one end and products removed at one or both ends. Points are experimental values and associated curves are calculations. Circles, 1.13 μmol/s injected at $x=1$ m and removed at $x=0$. Squares, 0.21 μmol/s injected at $x=1$ m and removed at $x=0$. Triangles, 3.36 μmol/s injected at $x=1$ m, 2.88 μmol/s removed at $x=1$ m, and 0.43 μmol/s removed at $x=0$.

creasing $\dot{N}$ in Eq. (7) is apparent: A reduction in the slope $|dn_H/dx|$ and an increase in the curvature of $n_H(x)$ as $\dot{N}$ is increased. These effects are analogous to the slope changes and curvature seen in $T_m(x)$ in thermoacoustic refrigeration by Reid et al.;[25] Reid and Swift,[26] and Reid[27] because $n_H \dot{N}$ and the terms with dn_H/dx in Eq. (7) here are mathematically identical in form to $\dot{m}c_p T_m$ and terms with dT_m/dx in the energy equation in the work of Reid, Swift, and coworkers. To the extent that the implicit n_H and x dependencies of all variables in Eq. (7) could be neglected, that equation would be of the form

$$\dot{N}_H = \alpha + \dot{N}n_H + \beta \frac{dn_H}{dx}, \qquad (18)$$

where α and β can be regarded as constants, so $n_H(x)$ would have an exponential curvature.

Figure 3(b) is less symmetrical than Fig. 3(a) because 20% of the feedstock mole flux is removed at $x=0$ and 80% at $x=1$ m in Fig. 3(b). The slopes are shallower and the curvatures greater on the right half of the figure than on the left half because $\dot{N}$ is greater on the right half, as discussed in the previous paragraph. The heavy mole fraction at $x=0.5$ m is different from $n_{H,\text{feedstock}}=0.50$, with the difference growing as the feedstock mole flux decreases and exerts weaker local influence on $n_H(x)$.

Figure 4 illustrates most clearly the sometimes counter-intuitive nature of the boundary condition for n_H expressed in Eq. (11). For the circles and squares, feedstock with $n_{H,\text{feedstock}}=0.50$ is injected at $x=1$ m and all of the product is removed at $x=0$. One might at first guess that n_H would be close to 0.50 at the feedstock end of the separation tube, but instead n_H is tied firmly to 0.50 at the *product* end of the separation tube. One way to interpret this result is to realize that, in steady state with only one feedstock capillary and one product capillary, what flows out of the product capillary must be exactly what flows in through the feedstock capillary, which has $n_{H,\text{feedstock}}=0.50$. In the context of Eqs. (9) and (10) there is only one term on the right and one on the left, so $n_{H,\text{product}}=n_{H,\text{feedstock}}$. In the context of Eq. (11), the product total mole flux and heavy mole flux must equal their values for the feedstock, so the product's n_H must also equal its value in the feedstock.

For the operating point of Fig. 3(b), in which 20% of $\dot{N}_{\text{feedstock}}$ flows out at $x=0$ and 80% at $x=1$ m, the corresponding global steady-state condition based on Eq. (10) shows that

$$0.20n_H(0) + 0.80n_H(1.0 \text{ m}) = n_{H,\text{feedstock}} = 0.50, \qquad (19)$$

which is indeed the case in Fig. 3(b). The triangles in Fig. 4 also illustrate Eq. (19) for one feedstock and two-product capillaries, though with different numerical values. In contrast to Fig. 3(b), in which the feedstock enters at the center of the separation tube and the two products are removed at the ends, the triangles in Fig. 4 represent a case in which the feedstock enters at one end and the two products are removed at the two ends.

Figure 5 shows how n_H in the two product streams at the ends of the separation tube varies with flow rate, when the feedstock is injected at the middle of the separation tube. Purities decrease with increasing flow rates, and the purity of one product can be enhanced by reducing the fraction of the feedstock that flows into it. For this separation tube running with a 3-kPa traveling wave, product flow rates of the order of 1 to 10 μmol/s cause a significant decrease in purities.

IV. CONCLUSIONS

Measurements confirm that the addition of the simple term $n_H \dot{N}$ to the thermoacoustic mixture-separation equation for the heavy mole flux $\dot{N}_H$ accounts for a nonzero total mole flux $\dot{N}$. The influence of $\dot{N}$ on $n_H(x)$ is similar to that of steady flow on $T_m(x)$ in the stack of a standing-wave refrigerator: As $\dot{N}$ is increased, the magnitude of the slope $|dn_H/dx|$ is reduced and $n_H(x)$ acquires significant curvature. Measurements also show that a boundary condition constraining the solution $n_H(x)$ is imposed wherever steady flow leaves the separation tube and enters a small-diameter capillary with negligible thermoacoustic characteristics of its

FIG. 5. Heavy mole fraction n_H at $x=0$ (points and curves above $n_H=0.5$) and at $x=1$ m (points and curves below $n_H=0.5$) as a function of product mole flux at $x=0$. Feedstock is injected at $x=0.5$ m and removed at $x=0$ and $x=1$ m. Points are experimental values and associated curves are calculations. Circles, feedstock flow equal to twice the $x=0$ product flow. Squares, feedstock flow equal to five times the $x=0$ product flow.

own. Flow out through such a capillary carries whatever mole fraction is present in the separation tube where the capillary is attached. No such boundary condition is imposed by feedstock entering the separation tube through such a capillary.

Capillaries suitable for practical continuous thermoacoustic mixture separation have been demonstrated. Such capillaries are small enough in diameter that they do not develop significant internal concentration gradients, nor do they consume significant acoustic power. Nevertheless they are large enough to carry significant steady flow without prohibitively large steady pressure drops.

ACKNOWLEDGMENTS

APPENDIX A: THERMOACOUSTIC MIXTURE SEPARATION IN CIRCULAR TUBES

The time-averaged mole flux of the heavy component when $dT_m/dx = 0$ is given by Eq. (7), where the three functions F_{trav}, F_{stand}, and F_{grad} depend on the properties of the gas and the radius of the tube. Analytical expressions for the three F's in the boundary-layer limit have been published previously.[1,3] Here we derive such expressions for circular tubes of arbitrarily small radius, starting from Eq. (5) and from Eqs. (47) and (48) of Ref. 3.

Unfortunately, Eq. (48) of Ref. 3 has a minus-sign typographical error; the correct expression is

$$
\dot{N}_{H,2} = \frac{1}{2}\frac{\gamma-1}{\gamma}\frac{k_T/\varepsilon}{R_{\text{univ}}T_m}\Re\left\{\frac{p_1\widetilde{U}_1}{1-\widetilde{f}_\nu}\left[C\left(\frac{\delta_\kappa^2}{\delta_{\kappa D}^2}-1\right)\langle h_{\kappa D}(1\right.\right.
$$
$$
\left.-\widetilde{h}_\nu)\rangle + (1-B-C)\left(\frac{\delta_\kappa^2}{\delta_{D\kappa}^2}-1\right)\langle h_{D\kappa}(1-\widetilde{h}_\nu)\rangle
$$
$$
\left.\left.-B\frac{\sigma-1}{\sigma}\langle h_\nu(1-\widetilde{h}_\nu)\rangle\right]\right\},
\tag{A1}
$$

where the last line here has the sign error fixed. B and C are given in Eqs. (33) and (34) of Ref. 3. Note that B is proportional to the concentration gradient and that C contains one term proportional to B and one term independent of the gradient. The $\langle\;\rangle$ denotes the average over the cross-sectional area, σ is the Prandtl number, the tilde denotes complex conjugation, and

$$
h_j = \frac{J_0[(i-1)r/\delta_j]}{J_0[(i-1)R/\delta_j]},
\tag{A2}
$$

$$
f_j = \frac{2J_1[(i-1)R/\delta_j]}{[(i-1)R/\delta_j]J_0[(i-1)R/\delta_j]},
\tag{A3}
$$

where J_0 and J_1 are the zeroth- and first-order Bessel functions, and j is ν, κD, or $D\kappa$. The coefficient ε and the penetration depths $\delta_{\kappa D}$ and $\delta_{D\kappa}$ are also defined in Ref. 3.

To evaluate averages of the form $\langle h_i(1-\widetilde{h}_j)\rangle$, one can use identity (11.3.20) from Ref. 28

$$
\int_0^z t^\nu J_{\nu-1}(t)dt = z^\nu J_\nu(z) \quad \text{for } \mathrm{Re}[\nu] > 0
\tag{A4}
$$

to verify that

$$
\langle h_j\rangle = f_j.
\tag{A5}
$$

Then one can use Ref. 28's Eqs. (9.1.40)

$$
\widetilde{J_\nu}(z) = J_\nu(\bar{z})
\tag{A6}
$$

and (11.3.29)

$$
\int^z\left\{(k^2-l^2)t-\frac{(\mu^2-\nu^2)}{t}\right\}\mathcal{C}_\mu(kt)\mathcal{D}_\nu(lt)dt
$$
$$
= z\{k\mathcal{C}_{\mu+1}(kz)\mathcal{D}_\nu(lz)-l\mathcal{C}_\mu(kz)\mathcal{D}_{\nu+1}(lz)\}-(\mu
$$
$$
-\nu)\mathcal{C}_\mu(kz)\mathcal{D}_\nu(lz),
\tag{A7}
$$

where $\mathcal{C}$ and $\mathcal{D}$ are any two Bessel functions, to compute

$$
\langle h_i\widetilde{h}_j\rangle = \frac{\delta_j^2}{\delta_i^2+\delta_j^2}f_i + \frac{\delta_i^2}{\delta_i^2+\delta_j^2}\widetilde{f}_j.
\tag{A8}
$$

Combining Eqs. (A5) and (A8) yields the result needed in Eq. (A1)

$$
\langle h_i(1-\widetilde{h}_j)\rangle = \frac{\delta_i^2}{\delta_i^2+\delta_j^2}(f_i-\widetilde{f}_j),
\tag{A9}
$$

and identification of expressions for the F's in Eq. (A1) is straightforward. The results are

$$
F_{\text{trav}} = -\frac{R}{\delta_\kappa}\,\mathrm{Re}\left[\frac{G}{(1-\widetilde{f}_\nu)}\right],
\tag{A10}
$$

$$
F_{\text{stand}} = \frac{R}{\delta_\kappa}\,\mathrm{Im}\left[\frac{G}{(1-\widetilde{f}_\nu)}\right],
\tag{A11}
$$

$$
F_{\text{grad}} = -\frac{R}{\delta_\kappa}\frac{1}{|1-f_\nu|^2}\frac{\sigma}{(1-\sigma)(1-\sigma L)-\varepsilon\sigma}
$$
$$
\times\mathrm{Im}\left[\frac{(\sigma-1)}{\sigma S}f_\nu\left(\frac{\delta_\kappa^2-\delta_{\kappa D}^2}{\delta_\nu^2+\delta_{\kappa D}^2}f_{\kappa D}-\frac{\delta_\kappa^2-\delta_{D\kappa}^2}{\delta_\nu^2+\delta_{D\kappa}^2}f_{D\kappa}\right.\right.
$$
$$
\left.\left.+\frac{(1+\sigma)LQ}{M}\widetilde{f}_\nu+S\right)+\varepsilon G\right],
\tag{A12}
$$

where R is the circle radius, $L = k/\rho_m c_p D_{12}$, and

$$
S = \left(\frac{\delta_\kappa^2}{\delta_{D\kappa}^2}-1\right)f_{D\kappa}-\left(\frac{\delta_\kappa^2}{\delta_{\kappa D}^2}-1\right)f_{\kappa D},
\tag{A13}
$$

$$
Q = \frac{\delta_{\kappa D}^2-\delta_{D\kappa}^2}{\delta_\kappa^2},
\tag{A14}
$$

$$
M = (1+\sigma)(1+\sigma L)+\varepsilon\sigma,
\tag{A15}
$$

$$G = \frac{\sigma L Q}{MS} f_{\kappa D} f_{D\kappa} + \frac{\tilde{f}_\nu}{S}\left(\frac{f_{\kappa D}}{1 + \delta_\nu^2/\delta_{D\kappa}^2} - \frac{f_{D\kappa}}{1 + \delta_\nu^2/\delta_{\kappa D}^2}\right). \quad \text{(A16)}$$

APPENDIX B: THERMOACOUSTICS IN CAPILLARIES

Ignoring the effects of steady flow, the wall-damped Helmholtz equation for pure gases is[10]

$$[1 + (\gamma - 1)f_\kappa]p_1 + \frac{a^2}{\omega^2}(1 - f_\nu)\frac{d^2p_1}{dx^2} = 0, \quad \text{(B1)}$$

with the viscous-dissipation function f_ν and the thermal-dissipation function f_κ in a circular tube of radius R given by

$$f_j = \frac{2J_1[(i-1)R/\delta_j]}{[(i-1)R/\delta_j]J_0[(i-1)R/\delta_j]}, \quad \text{(B2)}$$

where J_n is the nth-order Bessel function, $\delta_\nu = \sqrt{2\mu/\omega\rho_m}$ is the viscous penetration depth, and $\delta_\kappa = \sqrt{2k/\omega\rho_m c_p}$ is the thermal penetration depth, and j is either ν or κ. For small R, Eq. (B2) becomes

$$f_j \simeq 1 - \frac{i}{4}\frac{R^2}{\delta_j^2} - \frac{1}{12}\frac{R^4}{\delta_j^4} + O(R^6). \quad \text{(B3)}$$

Substitution of Eq. (B3) into Eq. (B1) yields, to lowest order in R, the Helmholz equation for a small circular tube

$$\left(1 - \frac{i}{4}\frac{\gamma-1}{\gamma}\frac{R^2}{\delta_\kappa^2}\right)p_1 + \frac{i}{4}\frac{R^2}{\delta_\nu^2}\frac{a^2}{\gamma\omega^2}\frac{d^2p_1}{dx^2} = 0. \quad \text{(B4)}$$

The solution $p_1(x)$ to Eq. (B4) is the sum of positive and negative complex exponentials of x. To conveniently satisfy boundary conditions at the ends of the capillary, the solution can be written in either of two forms

$$p_1 = p_{1,0}e^{-\alpha x}, \quad \text{infinite length}, \quad \text{(B5)}$$

$$p_1 = p_{1,0}\frac{\sinh \alpha(\Delta x - x)}{\sinh \alpha\Delta x}, \quad \text{finite length}, \quad \text{(B6)}$$

for capillaries driven at $x=0$ with a complex pressure amplitude $p_{1,0}$. The first form is for a capillary extending from $x=0$ to infinity, and the second form is for a capillary of length Δx with $p_1=0$ at $x=\Delta x$. Substitution of either form into Eq. (B4) and use of the general identity $a=\lambda f$ relating the sound speed a, the wavelength λ, and the frequency f shows that

$$\alpha^{-1} = \frac{1-i}{\sqrt{2}}\frac{1}{4\pi\sqrt{\gamma}}\frac{R}{\delta_\nu}\lambda \quad \text{(B7)}$$

to lowest order in R. The volume flow rate U_1 is obtained from Eqs. (B5)–(B7) via Eq. (2) taken to the small-R limit

$$U_1(x) = \frac{i\pi R^2(1-f_\nu)}{\omega\rho_m}\frac{dp_1}{dx} \quad \text{(B8)}$$

$$\simeq -\frac{\pi R^4}{8\mu}\frac{dp_1}{dx}, \quad \text{(B9)}$$

which yields the expected Poiseuille fourth-power dependence on radius.[20]

For a gas mixture, f_κ in Eq. (B1) should be replaced by a linear combination of $f_{\kappa D}$ and $f_{D\kappa}$, which are defined in Ref. 3. However, the absence of δ_κ from Eqs. (B5)–(B7) shows that this detail is irrelevant for the determination of α in the present problem.

Table I displays some numerical values for the gas, the microcapillaries, the feedstock/product capillaries, and the separation tube. In the microcapillaries, $|e^{-\alpha\Delta x}| = 2\times 10^{-6}$, so Eq. (B5) is appropriate. The feedstock/product capillaries have $|e^{-\alpha\Delta x}| = 0.06$, so Eq. (B6) would be required for high accuracy, but for simplicity it is not used here.

To estimate the end-to-end concentration difference that develops in a capillary, we begin by considering Eq. (7). The two terms with dn_H/dx typically subtract from the $|p_1||U_1|$ term and are smaller than it. Thus, $\dot{N}_H - n_H\dot{N}$ will never be much larger than the $|p_1||U_1|$ term, and $\dot{N}_H - n_H\dot{N}$ can be neglected when estimating the order of magnitude of dn_H/dx in the capillary. In small enough tubes or at low enough amplitudes, the third term in Eq. (7) can be neglected in comparison to the fourth term, because their ratio

$$\frac{\text{third term}}{\text{fourth term}} = \frac{1}{16\pi}\frac{L}{\gamma\sigma}\frac{R\lambda^2}{\delta_\kappa^3}F_{\text{grad}}\frac{|p_{1,0}|^2}{p_m^2}e^{-2\,\text{Re}[\alpha]x} \quad \text{(B10)}$$

is proportional to $R^4|p_{1,0}|^2$. (Note that F_{grad} is proportional to R^3 for small R.) To obtain Eq. (B10) and subsequent results in this section, it is necessary to use Eqs. (B9), (B7), and (B5), the definitions $\sigma = \delta_\nu^2/\delta_\kappa^2$ and $L = \delta_\kappa^2/\delta_D^2$ given in Ref. 3, and the ideal-gas identities $a^2 = \gamma R_{\text{univ}}T_m/m_{\text{avg}}$ and $p_m = \rho_m R_{\text{univ}}T_m/m_{\text{avg}}$. Based on the numerical values in Table I for Eq. (B10), we neglect the third term in Eq. (7).

With these approximations, Eq. (7) becomes

$$0 \simeq \frac{\delta_\kappa}{2R}\frac{\gamma-1}{\gamma}\frac{k_T}{R_{\text{univ}}T_m}(F_{\text{trav}}\cos\theta + F_{\text{stand}}\sin\theta)|p_1||U_1|$$

$$- N\pi R^2 D_{12}\frac{dn_H}{dx} \quad \text{(B11)}$$

in a capillary. Solving this equation for dn_H/dx and integrating along the length of the capillary, using Eqs. (B5)–(B7) as necessary, with $\theta = 3\pi/4$ from Eqs. (B5) and (B9), and using the small-R limits of F_{trav} and F_{stand}

$$F_{\text{trav}} \propto R^5, \quad \text{(B12)}$$

$$F_{\text{stand}} \simeq \frac{1}{12}\frac{R^3}{\delta_\kappa^3}, \quad \text{(B13)}$$

yields finally

$$\Delta n_H \simeq \frac{1}{384\pi^2}\frac{(\gamma-1)L}{\gamma^2\sigma}k_T\frac{\lambda^2R^4}{\delta_\kappa^6}\frac{|p_{1,0}|^2}{p_m^2} \quad \text{(B14)}$$

for infinite length. Numerical values are given in Table I.

The acoustic power consumed by each capillary can be calculated with

$$\dot{E}_2(0) = \frac{1}{2}\text{Re}[p_1(0)\tilde{U}_1(0)]. \quad \text{(B15)}$$

Substituting Eq. (B9) into Eq. (B15) yields

G. W. Swift and D. A. Geller: Continuous thermoacoustic mixture separation

$$\dot{E}_2 = \sqrt{\frac{\gamma}{32}} \frac{|p_{1,0}|^2}{\rho_m a} \pi R^2 \frac{R}{\delta_\nu} \tag{B16}$$

for infinite length. Numerical values are given in Table I.

[1] G. W. Swift and P. S. Spoor, "Thermal diffusion and mixture separation in the acoustic boundary layer," J. Acoust. Soc. Am. **106**, 1794–1800 (1999); **107**, 2299(E) (2000); **109**, 1261 (2001).

[2] P. S. Spoor and G. W. Swift, "Thermoacoustic separation of a He–Ar mixture," Phys. Rev. Lett. **85**, 1646–1649 (2000).

[3] D. A. Geller and G. W. Swift, "Saturation of thermoacoustic mixture separation," J. Acoust. Soc. Am. **111**, 1675–1684 (2002).

[4] D. A. Geller and G. W. Swift, "Thermodynamic efficiency of thermoacoustic mixture separation," J. Acoust. Soc. Am. **112**, 504–510 (2002).

[5] A computer animation of this process can be obtained by downloading MixSepZip.exe from www.lanl.gov/thermoacoustics/movies.html

[6] D. A. Geller and G. W. Swift, "Thermoacoustic enrichment of the isotopes of neon," J. Acoust. Soc. Am. **115**, 2059–2070 (2004).

[7] R. W. Rousseau, ed., *Handbook of Separation Process Technology* (Wiley, New York, 1987).

[8] D. M. Ruthven, ed., *Encyclopedia of Separation Technology* (Wiley, New York, 1997).

[9] S. Villani, *Isotope Separation* (American Nuclear Society, La Grange Park, IL, 1976).

[10] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America Publications, Sewickley, PA, 2002).

[11] W. C. Ward and G. W. Swift, "Design environment for low amplitude thermoacoustic engines (DeltaE)," J. Acoust. Soc. Am. **95**, 3671–3672 (1994). Software and user's guide available either from the Los Alamos thermoacoustics web site www.lanl.gov/thermoacoustics/or from the Energy Science and Technology Software Center, US Department of Energy, Oak Ridge, Tennessee.

[12] MKS Baratron Type 390HA, www.mksinst.com

[13] Endevco, San Juan Capistrano, CA, www.endevco.com

[14] Model SR830, Stanford Research Systems, Sunnyvale CA, www.srsys.com

[15] Polymicro Technologies LLC, www.polymicro.com

[16] RGA 100, Stanford Research Systems, Sunnyvale CA, www.srsys.com

[17] Turbo-V 70D, Varian Inc. Vacuum Technologies, www.varianinc.com

[18] Dry Scroll Vacuum Pump SH-110, Varian Inc. Vacuum Technologies, www.varianinc.com

[19] M. G. Rao and C. Dong, "Evaluation of low cost residual gas analyzers for ultrahigh vacuum applications," J. Vac. Sci. Technol. A **15**, 1312–1318 (1997).

[20] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1982) Eq. (17.10).

[21] Model S, Swagelok Inc., www.swagelok.com

[22] Model PRG101, Omega Engineering, Stamford CT, www.omega.com

[23] M. Escorne and A. Mauger, "Temperature control for liquid-helium cryostats below 4.2 K," Rev. Sci. Instrum. **54**, 1693–1696 (1983).

[24] B. L. Smith and G. W. Swift, "Measuring second-order time-average pressure," J. Acoust. Soc. Am. **110**, 717–723 (2001).

[25] R. S. Reid, W. C. Ward, and G. W. Swift, "Cyclic thermodynamics with open flow," Phys. Rev. Lett. **80**, 4617–4620 (1998).

[26] R. S. Reid and G. W. Swift, "Experiments with a flow-through thermoacoustic refrigerator," J. Acoust. Soc. Am. **108**, 2835–2842 (2000).

[27] R. S. Reid, "Open cycle thermoacoustics," Ph.D. thesis, Georgia Institute of Technology, School of Mechanical Engineering, 1999.

[28] M. Abramowitz and I. E. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (National Bureau of Standards, US Government Printing Office, Washington DC, 1964).

Properties of an electrostatic transducer

A. P. Medley,[a] D. R. Billson, D. A. Hutchins, and L. A. J. Davis
School of Engineering, University of Warwick, Coventry, CV4 7AL United Kingdom

A thin and flexible transducer design for use in air is described. It is fabricated from three or more layers of material to form an electrostatic device with many interesting properties. The new acoustic source has an excellent high-frequency response and can be used as an acoustic radiator over a wide range of frequencies, in both the audible and ultrasonic frequency ranges. The frequency response and directivity are both modeled and compared to theory, where it is demonstrated that the device operates in a manner consistent with a plane piston source. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2357702]

I. INTRODUCTION

This paper introduces an electrostatic loudspeaker that is both extremely thin and highly flexible. The device can operate in both the audible and ultrasonic frequency ranges. As a consequence, this new design has many interesting and varied potential applications. For example, the transducer can be produced in the form of a large area planar device, which can be mounted directly onto a wall to provide sound reproduction in a home environment without requiring a bulky enclosure. Furthermore, such a device would be particularly suitable for use in applications where space is at a premium, or where an unconventional or adjustable curvature is required.

The moving coil loudspeaker has existed for many decades and is still by the far the most common method for sound transduction. Although a number of different technologies have been developed, the electrostatic principle has been used in several designs of high performance loudspeakers. Wente[1] in 1917 was the first to describe a device using electrostatic principles. This transducer, which was called a condenser transmitter and was originally used as a microphone for measuring absolute sound intensity, utilized a thin stretched steel membrane, separated from a conducting backplate by a 25-μm-thick air gap. Through experimentation it was shown that such a device had a reasonably uniform sensitivity at frequencies of up to 10 kHz, and an undamped resonant frequency of approximately 17 kHz. Further development of electrostatic loudspeakers was hampered by a lack of suitable materials for the vibrating membrane. Early patents show that gelatin, silk, rubber, and other materials were considered,[2] but the development of thin polymer membranes such as Mylar made the modern designs possible. The lack of bass response from electrostatic speakers means that they are often incorporated into hybrid systems, whereby a standard moving-coil speaker is used for low frequencies.[3] Other developments include the use of multiple electrodes on the moving diaphragms or membranes, so that the proportion of radiating area used can be split into different frequency bands via a crossover network.[4,5]

In recent years there has been increased interest in the use of electrostatic devices for acoustic generation at ultrasonic frequencies. This is because the efficiency of conventional piezoelectric transducers is affected by the large impedance mismatch between the piezoelectric element and air. This can be improved by several methods, such as attaching a metal membrane to the element to act as a vibrating membrane,[6] by using impedance matching layers,[7] or by using piezocomposites to lower the acoustic impedance of the active material.[8] An alternative approach to that of a piezoelectric device is to use a capacitive design. For example, Kuhl *et al.*[9] produced transducers capable of operating in the 50–100-kHz range, and this was the basis upon which all modern capacitive transducers are designed. They basically rely on the use of a flexible electroded membrane, separated by an air gap from a conducting rigid backplate. The response of the device is dominated by the interaction between the membrane and the trapped air, and various models have been proposed to optimize this.[10,11]

It is the aim of this paper to describe a new type of device, which differs from the two basic designs above in that it is comprised of multiple layers, which are all thin and flexible. This paper begins with a description of the basic structure of the transducer, which is then modeled both numerically and analytically so as to predict the frequency response of the device in terms of the layered construction. This is then followed by a series of experimental measurements, using a testing arrangement in which the properties of the speaker can be easily altered. This enabled the results of experimental device responses and modeling to be compared, so as to enable the operation of the device to be fully understood. Finally, the pressure fields radiated from the acoustic transducer have been obtained experimentally over a range of frequencies, and the results compared to a theory, allowing a greater understanding of the behavior of the transducer to be developed.

II. THEORETICAL ANALYSIS

The new acoustic transducer consists of three layers of thin, flexible material, which are sandwiched together to form the device. This is shown schematically in Fig. 1, where it can be seen that a typical device has an insulating

---

[a]Electronic mail: a.p.medley@warwick.ac.uk

FIG. 1. Schematic representation of the acoustic transducer.

TABLE I. Material properties of the three media relevant to the initial designs shown in Fig. 2.

Layer	Density ρ_i (kg/m³)	Wave velocity (m/s)	
		Longitudinal v_{li}	Shear v_{si}
Aluminum	2700	6250	3080
Air	1.2	343.3	⋯
Paper	780	400	200
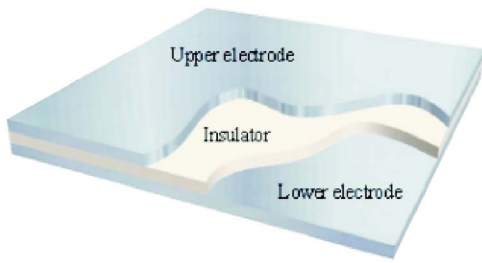
layer between two electrically conducting layers. As with conventional electrostatic devices, a dc bias voltage is applied between the front and rear electrodes, attracting the flexible layers to each other. The electrodes can be thin metallic layers (e.g., aluminum), or metallized polymer films (e.g., Mylar). Because the individual layers are so close together, the application of a modest ac voltage signal creates a large electric field, which is highly efficient at moving the foils in the thickness direction, generating sound. However, the device will change characteristics depending on the materials used for each layer in terms of density and thickness, and the thickness of any air gaps between them.

There are many examples of multilayered structures that have been analyzed in terms of their acoustic response, included adhesively bonded joints.[12–14] Theoretical models can be used to predict the response of materials, and to search for defects by looking at the effects these have on the vibrational response.[14] One such method is the matrix transfer approach, as detailed by Brekhovskikh[15] and used by Bresse et al.[16] This method calculates the transmission coefficient as a function of frequency when the structure is excited by an infinite sinusoidal plane wave in air at an arbitrary angle to the structure's surface. All waves in the structure will have a common propagation factor. Therefore, by using the boundary conditions that exist between each layer, a propagation matrix can be established such that the coefficients of reflection and transmission of the top layer can be related to those of the bottom layer of the structure. Thus, a description of the response of the device can be obtained. The original algorithm was originally designed for liquid-coupled and contacting experiments. However, it is sufficiently general to allow air to be defined as the coupling medium and for any number of layers to be modeled. While this through-transmission model does not describe the loudspeaker as a source, it is a useful method for predicting the expected frequency response of such a device in air. The model assumes normal incidence of plane waves, whereas the experimental electrostatic forces would give a uniform excitation of the membranes; however, the mechanical response would be expected to be similar in both cases in terms of the frequency response.

Consider the typical structure shown in Fig. 1, constructed with two aluminum foil outer layers, an inner paper layer, and a 1-mm air gap between each. Table I gives details of the properties of each layer used in the calculation. Figure 2 illustrates the three structures modeled initially, and the corresponding predictions of the transmission frequency response are presented in Fig. 3.

It can be seen that the transmission coefficient approaches unity at low frequencies. This is because the structure is thin compared to the acoustic wavelength. As the frequency increases, the transmission coefficient decreases rapidly towards zero, since the acoustic impedance of the structure becomes much larger than that of the surrounding air. The transmission coefficient then exhibits a series of maxima, corresponding to through-thickness resonances of the layers, which make up the overall structure. These can clearly be seen in Fig. 3(a), in which two resonances occur, the first at 10 kHz and the second at 14.4 kHz. In Fig. 3(c) it can be seen that two resonances are closer together and have tended to merge to become a wider peak in the predicted response.

It is informative to also analyze the structure in terms of a simple mathematical model, as this allows us to examine the likely behavior and motion of the layers within the device. Using the assumption the membranes can be considered to move as a simple frictionless piston resonator, the speaker can be modeled as a simple mass spring system. The air located between the solid layers is compressible and of very low mass. It can thus be considered to behave like a massless spring when compared to the membranes (see Fig. 4).

The equation of motion for the system can be expressed as

$$m_1 \ddot{x}_1 + k_1 x_1 - k_1 x_2 = 0,$$
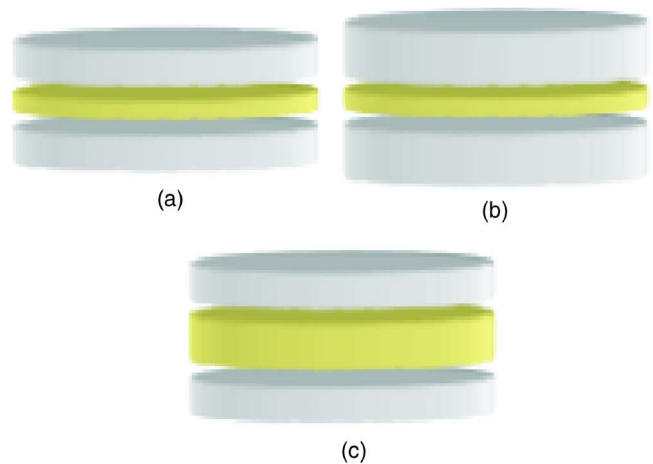
FIG. 2. Schematic diagram of three typical structures. All devices are constructed from two outer layers of aluminum and an inner layer of paper, separated by 1-mm-thick air gaps. (a) 14 μm aluminum and 90 μm paper; (b) as in (a) but with thicker aluminum (28 μm); (c) as in (a) but with thicker paper (180 μm). The properties of each layer are defined in Table I.

(a)

(b)

(c)
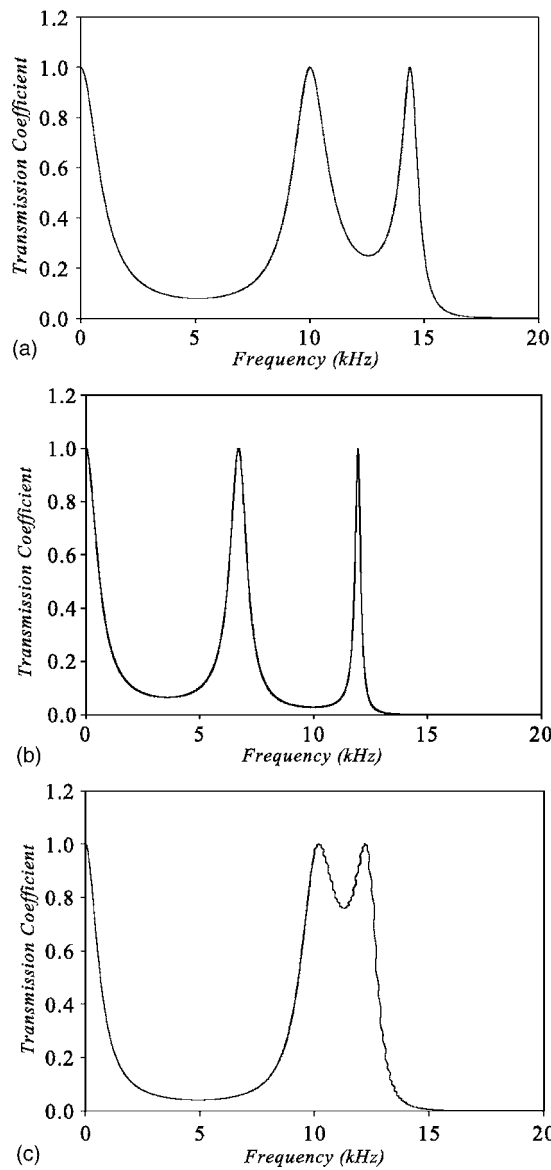
FIG. 3. (a)–(c) Calculated values of transmission coefficient as a function of frequency for the corresponding multilayered structures shown in Fig. 2 and with the properties of Table I.

$$m_2\ddot{x}_2 - k_1 x_1 + (k_1 + k_2)x_2 - k_2 x_3 = 0,$$

$$m_3\ddot{x}_3 - k_2 x_2 + k_2 x_3 = 0, \qquad (1)$$

where m_1, m_2, and m_3 are the upper electrode, insulator, and lower electrode membrane masses, respectively, k_1 and k_2 are the spring constants, and x_1, x_2, and x_3 are the displacement of the masses. It should be noted that for simplicity neither
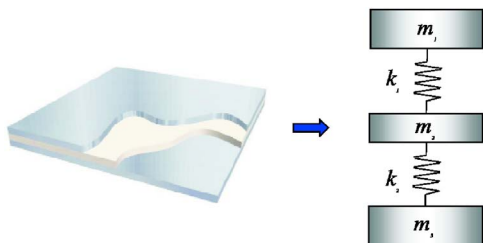
FIG. 4. Modeling assumption for the multilayered acoustic transducer.
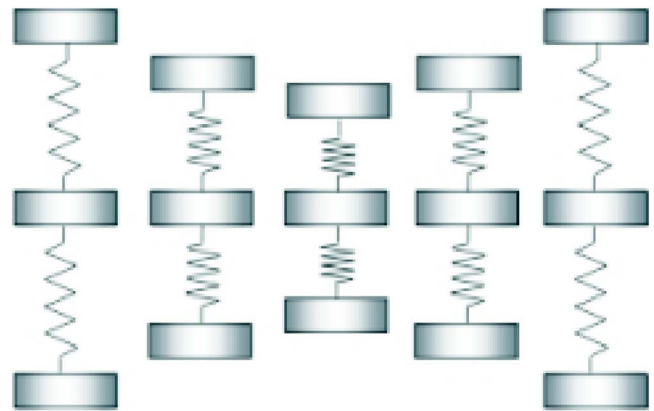
FIG. 5. Mode shape for $\lambda = k/m_1$.

the electrostatic force nor negative stiffness caused by the application of a bias voltage has been included. Assuming that the spring constant is the same for both springs, and that the two outer masses are the same, these equations can be rewritten in matrix form as

$$\begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_1 \end{bmatrix}\begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \ddot{x}_3 \end{Bmatrix} + \begin{bmatrix} k & -k & 0 \\ -k & 2k & -k \\ 0 & -k & k \end{bmatrix}\begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = 0. \qquad (2)$$

Solving this matrix equation enables the eigenvalues of the system to be evaluated:

$$\lambda = 0,$$

$$\lambda = k/m_1,$$

$$\lambda = (2m_1 k + m_2 k)/(m_1 m_2). \qquad (3)$$

The null value eigenvalue corresponds to a rigid body translation, which would not be observed upon application of an ac voltage, thus the actual number of resonant modes is one less than the number of degrees of freedom for the system. Therefore the corresponding eigenvectors of the system are

$$\{-1 \ 0 \ 1\},$$

$$\{1 \ -2m_1/m_2 \ 1\}. \qquad (4)$$

These show that for $\lambda = k/m_1$, as shown in Fig. 5, the two outer masses move in anti-phase, while the central mass remains fixed. For $\lambda = (2m_1 k + m_2 k)/(m_1 m_2)$, as shown in Fig. 6, the two outer masses move in phase with one another and with equal amplitude, such that the distance separating them does not change, while the central mass moves in anti-phase to the movement of both the upper and lower masses with double the amplitude, such that linear momentum is conserved.

The resonant frequencies, f_n, of a simple mass spring system are well known, and for the two cases are given by

$$f_1 = 1/2\pi\sqrt{k/m_1} \quad \text{and} \quad f_2 = 1/2\pi\sqrt{2k/m_2 + k/m_1}. \qquad (5)$$

Therefore, in order to find the resonant frequencies of the system, it is necessary to evaluate the two ratios k/m_1 and k/m_2. Within this model, the spring actually represents the
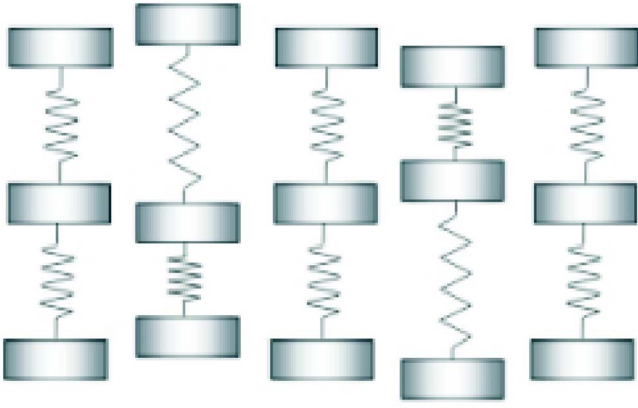
FIG. 6. Mode shape for $\lambda=(2m_1k+m_2k)/(m_1m_2)$; a fixed distance is maintained between the two outer masses.

air trapped between the layers. Assuming adiabatic compression results in the following expression for the stiffness, k_a, of the air spring:

$$k_a = \gamma PA/d_a, \tag{6}$$

where γ is the adiabatic constant, P is the pressure of the gas between the layers, A is the membrane area, and d_a is the thickness of the air gap between the layers of the structure when the system is at rest. Substituting this equation into Eq. (6) and letting $m_1=\rho_m d_m A$ gives

$$f = 1/2\pi\sqrt{\gamma P/d_a d_m \rho_m} \tag{7}$$

where ρ_m is the density of the membrane and d_m is the thickness of the conducting membrane. The equation above dictates that the resonant frequency increases with P, but decreases if d_a, d_m, or ρ_m increases. Note that the use of a one-dimensional model dictates that the result is independent of surface area (A).

The model so far neglects any mechanical properties of the vibrating membrane that can be assumed to influence the behavior of the system. This, however, can be rectified by considering the frequency response of a flat plate, the formulas for which are known.[17] As the membranes are not placed under significant tension (a factor that can affect platelike behavior in suspended systems), the tension can be ignored. Hence, the resonant frequency of a flat plate whose edges are fixed is given as

$$f = (K/2\pi)\sqrt{Dg/ma^4}, \tag{8}$$

where K is a ratio that is dependent upon the length of the plate edges (for a square plate $K=36$), m is the mass of the membrane, a is the length of a side, and g is the acceleration due to gravity. D is the flexural rigidity per unit length and can be equated to

$$D = Et^3/12(1-\nu^2) \tag{9}$$

where E is the modulus of elasticity, t is the thickness of the membrane, and ν is Poisson's ratio. By substituting $f=1/2\pi\sqrt{k/m}$ into $f=K/2\pi\sqrt{Dg/ma^4}$ the spring constant of the plate model can be evaluated as being equal to

$$k_p = K^2Dg/a^4. \tag{10}$$

In order to substitute this into the developed resonant frequency expression the basic model must be adapted to include a second "plate" spring to account for the mechanical properties of the device. Taking into account the added spring, the equations for the resonant frequency of the system become

$$f_1 = 1/2\pi\sqrt{(k_p+k_a)/m_1},$$

$$f_2 = 1/2\pi\sqrt{(2(k_p+k_a)/m_2) + ((k_p+k_a)/m_1)}. \tag{11}$$

Compliance is a measure of how easily an assembly will move when an electrical signal is applied to it. Mathematically it is the inverse of stiffness and can be calculated by

$$C = 1/k. \tag{12}$$

Therefore, substituting the values for C and setting $m_1 = \rho_{m_1}d_{m_1}A$ and $m_2=\rho_{m_2}d_{m_2}A$, where A is the area of the transducer, the resonant frequency equations of the system become

$$f_1 = \frac{1}{2\pi}\left[\frac{\gamma P}{d_a\rho_{m_1}d_{m_1}} + \frac{K^2Dg}{\rho_{m_1}d_{m_1}Aa^4}\right]^{\frac{1}{2}},$$

$$f_2 = \frac{1}{2\pi}\left[\left(\frac{2\gamma P}{d_a\rho_{m_2}d_{m_2}} + \frac{2K^2Dg}{\rho_{m_2}d_{m_2}Aa^4}\right) + \left(\frac{\gamma P}{d_a\rho_{m_1}d_{m_1}} + \frac{K^2Dg}{\rho_{m_1}d_{m_1}Aa^4}\right)\right]^{\frac{1}{2}}, \tag{13}$$

where the double subscript notation indicates which mass of the system the variable is attributed to. It would be more use if the equation could be related to the speed of sound in a gas, using

$$c^2 = \gamma RT \quad \text{and} \quad P = \rho_a RT. \tag{14}$$

Substituting these equations into the resonant frequency equation and cancelling by RT gives

$$f_1 = \frac{1}{2\pi}\left[\frac{c^2\rho_a}{d_a\rho_{m_1}d_{m_1}} + \frac{K^2Dg}{\rho_{m_1}d_{m_1}Aa^4}\right]^{\frac{1}{2}}, \tag{15}$$

$$f_2 = \frac{1}{2\pi}\left[\left(\frac{2\rho_a c^2}{d_a\rho_{m_2}d_{m_2}} + \frac{2K^2Dg}{\rho_{m_2}d_{m_2}Aa^4}\right) + \left(\frac{c^2\rho_a}{d_a\rho_{m_1}d_{m_1}} + \frac{K^2Dg}{\rho_{m_1}d_{m_1}Aa^4}\right)\right]^{\frac{1}{2}}. \tag{16}$$

It is interesting to note that the modeling procedure used can be extended to include any structure and combination of layers. Equations (15) and (16) indicate that the dominant factors that determine the resonant frequency of the structure are the density, speed of sound, and thickness of the air separating the membranes, and the density and thickness of the membrane material.

To assess the validity of the model against the results obtained from the numerical algorithm, numerical results were calculated using the material properties given in Table I, for the multilayered structures defined in Fig. 2, and plotted as vertical dashed lines in Figs. 7(a) and 7(b) to illustrate

FIG. 9. Apparatus used to allow controlled variations in air gap between three chosen solid layers in an electrostatic device.
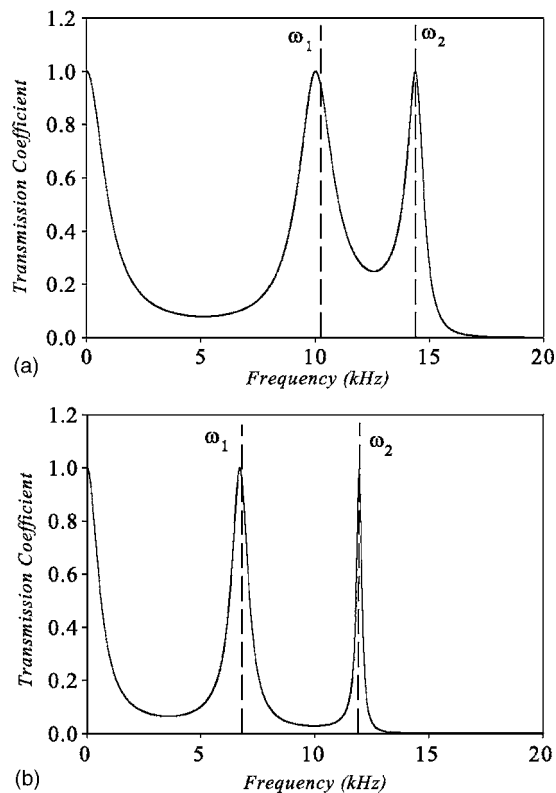
FIG. 7. The calculated response of the structure shown earlier in Fig. 2(a) for both the numerical model (solid line) and the analytical model (dotted line).

the single values of the predicted resonant frequencies f_1 and f_2. The predictions of the matrix transfer method given earlier in Fig. 3 are shown as the solid curves. There is a very good agreement between the analytical model and the response predicted by the numerical algorithm.

The analytical model also indicates that the dimensions of the device are an important factor in determining the resonant frequency. The inclusion of a second "plate" spring implies that as the lateral dimensions of the device are reduced, the mechanical properties of the plate dominate the response of the structure. This is illustrated for the particular case of the first resonance, f_1, in Fig. 8 for the structure shown in Fig. 2(a). This demonstrates that for wide devices, the re-
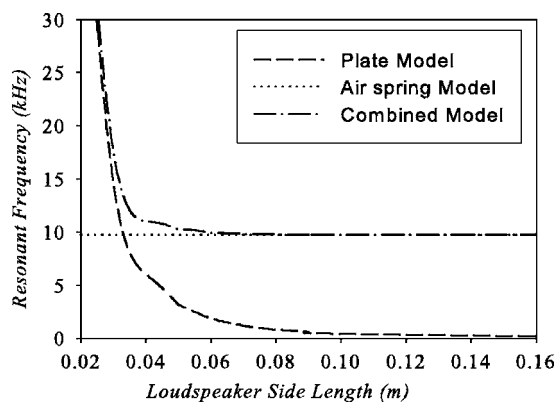
FIG. 8. Theoretical predictions for f_1, showing the individual contributions of the simple air-spring and plate models, and the result of combining them to get an overall response.

sponse is dominated by the air-spring model, but as the width of the device decreases, the plate response becomes more dominant and causes the predicted value of f_1 to increase dramatically. Note that for devices greater than 0.06 m in width, the response is dominated by the air-spring model.

Note that the above analytical model does not include damping. This can be considered by the addition of a damper to the air-spring system.[18] The new resonant frequency f_{1D} that results from damping is defined in terms of the bandwidth $\Delta\omega$ of the resonant peak and its value, as shown in Eq. (17), where an additional term is evident when compared to Eq. (15):

$$f_{1D} = \frac{1}{2\pi} \sqrt{\frac{c^2 \rho_a}{d_a \rho_{m_1} d_{m_1}} + \frac{K^2 Dg}{\rho_{m_1} d_{m_1} A a^4}} \sqrt{1 - \left(\frac{\Delta\omega}{2f_1}\right)^2}. \quad (17)$$

This tends to result in a small reduction in the predicted resonant frequency. The value of $\Delta\omega$ is difficult to predict theoretically and is thus determined experimentally. As a consequence, any alteration in the structure of the transducer that affects the resonance will also affect the value of the damping factor used.

III. EXPERIMENTS WITH A PROTOTYPE DEVICE

An apparatus was constructed to be able to test device configurations with controllable air gaps and layer configurations. The membranes, which measured 100 mm in diameter, were suspended on individual concentric rings, each of which was supported from the base by three springs in a trihedral arrangement. Three micrometers were attached to a top stationary ring, and these were used to push against the springs, thereby enabling the distance between the various membranes making up the device to be accurately controlled, as seen in the photograph of Fig. 9. This arrangement also allowed for the material used in the construction of the transducer to be easily altered.

Figure 10 is a schematic diagram of the apparatus used to measure the output from the device. The transducer was placed within an anechoic chamber and driven with a continuous sinusoidal waveform with a peak-to-peak voltage of
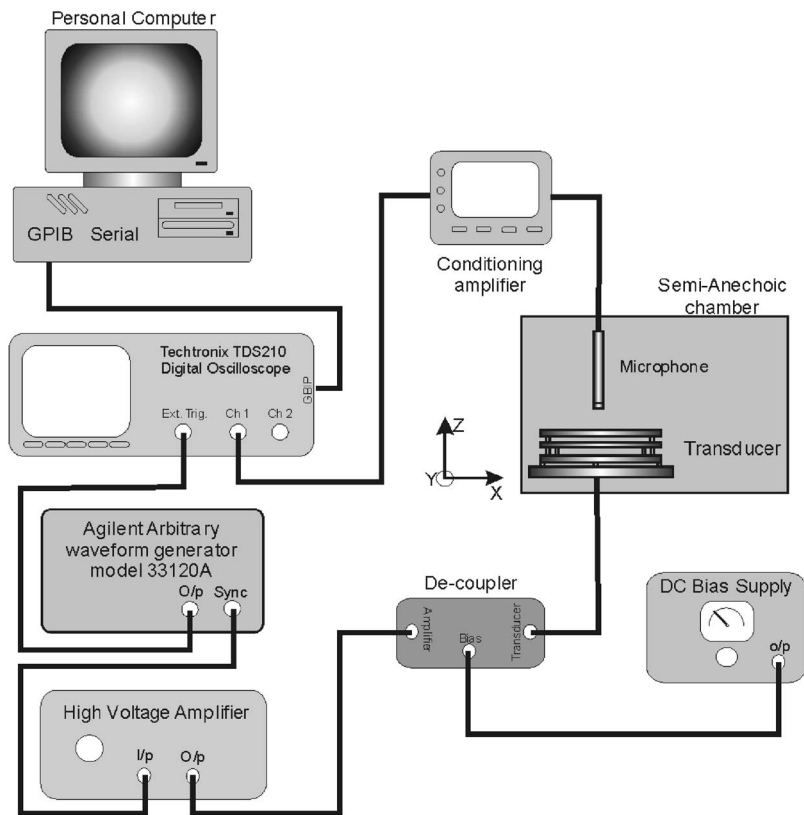
FIG. 10. Schematic diagram of the experimental arrangement used in determining the frequency response of the device shown in Fig. 9.

up to 230 V from an Agilent 33120A arbitrary waveform generator plus voltage amplifier, creating an electric field approximately equal to 2.5 MV m^{-1}. An additional 300-V dc bias was applied across the membranes of the transducer, to allow the drive voltage swing to remain undistorted, resulting in an ac to dc ratio of approximately 2/3, and experimentally harmonic distortion was not observed. The bias is necessary in electrostatic devices as the drive forces are always attractive; lack of a dc bias would lead to frequency doubling effects. The resulting acoustical output was measured with a Brüel and Kjær 4138 eighth-inch microphone, in conjunction with a Brüel and Kjær 2690 conditioning amplifier. The microphone had a calibrated frequency response of between 6.5 Hz and 140 kHz, and a maximum sensitivity of 1 mV/Pa. A preamplifier was incorporated into the slender body of the microphone; this had a 0-dB gain and a flat frequency response over the range 1 Hz to 200 kHz. The preamplifier and conditioning amplifier also both supported TEDS (transducer electronic data sheet), a plug and play data acquisition technology that automates the setup and configuration of the microphone with the conditioning amplifier. To automate the whole procedure a Labview™ virtual instrument was written, which enabled the transducer to be driven with a defined waveform over the frequency range 50 Hz to 20 kHz. The driving and measured waveforms were displayed on a Tektronics TDS210 digital oscilloscope and the waveform stored on a PC for future analysis.

A set of experiments was now performed in which the distance between the membranes of the prototype transducer was adjusted and the response from the microphone recorded on-axis at a distance of 10 mm from the front face of the device. Figure 11(a) shows a typical frequency response

from a 100-mm-diam transducer obtained experimentally, with the prediction of the numerical model reproduced alongside for comparison. The device had two 14-μm-thick outer aluminum layers and a 90-μm-thick paper central layer, with symmetrical air gaps of 1.75 mm. It can be seen that two resonant peaks are visible in the experimental data at approximately the same values as that predicted by theory. The large response at low frequencies is not observed experimentally because this would correspond to a rigid body translation. The model looks at the transmission coefficient of the device in air, which is comprised of two effects: the response of the device and the transmission through it. As explained earlier, the transmission coefficient would be predicted to approach unity for a thin device, whereas experimentally this is not measured, hence the differences at low frequencies. In addition, the experimental peak at approximately 3.6 kHz, which is not present in the theoretical predictions, is thought to be due to a standing wave between the transducer front surface and the microphone. Figure 11(b) shows the same comparison for a thicker (20 μm) membrane but a smaller air gap of 1 mm. Now the two resonant peaks are further apart in value, but shifted to higher frequencies.

Figure 12 depicts the predictions of both mathematical models and experiment for the value of the first resonant frequency (f_1), as a function of changes in (a) air gap for a constant membrane thickness and (b) membrane thickness for a constant air gap. Models and experiment agree fairly well, apart from small air gaps (<1 mm) at low frequencies, where the numerical solution deviates from other values as explained earlier. The experimental results generally show resonances slightly lower than predicted due to mass loading
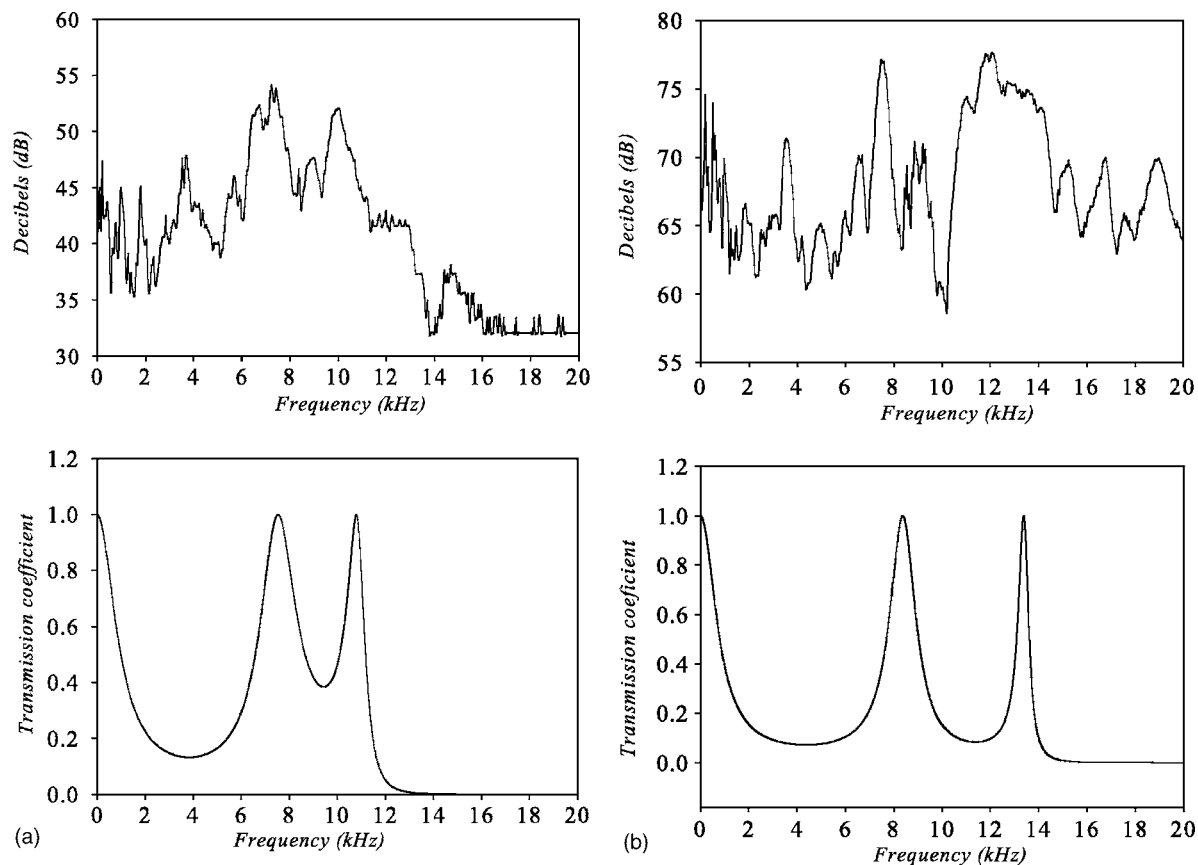
FIG. 11. Comparison of experimental frequency response (left) and that predicted numerically (right). (a) Device with 14-μm-thick outer aluminum layers, a 90-μm-thick paper central layer, with symmetrical air gaps of 1.75 mm between each layer. (b) A similar device but with 20-μm-thick aluminum and an air gap of 1 mm. In both instances the transducer had a diameter of 100 mm.

effects, which is a result of the "masses air spring" assumption used in the model. As expected, larger membrane thicknesses and bigger air gaps lead to a reduction in the expected fundamental resonant frequency, enabling the ultimate design goals of low fundamental resonance and wide frequency bandwidth of the device to be achieved.

Experiments were also performed in which the thickness of the central paper membrane was changed. From theoretical considerations, as predicted by Eqs. (15) and (16), this should not affect the value of f_1 [Eq. (15)], a symmetric vibration about this center layer, but should cause f_2 to change via Eq. (16). This was indeed observed experimentally.

During modeling of the transducer it was assumed that the membranes of the device could be considered to move as a simple piston resonator. While a comparison of the results obtained both experimentally and theoretically would seem to support this assumption, a series of experiments has been carried out to ascertain whether the assumption was correct. Using an optical detection system to determine the surface movement of the loudspeaker it was found that the device deviated from the plane piston response by 8%, further validating the initial assumption.

From the above, it can be seen that that the adjustable three-layer prototype device behaved in a manner that was consistent with theoretical predictions. The two resonances predicted by both numerical and analytical solutions were present in experimental data taken on-axis. One of the reso-

nances (at frequency f_1) was a symmetrical vibration about the central layer, whereas the second (at frequency f_2) was an asymmetric vibration with all three layers in motion.

IV. RADIATED FIELDS OF A PRACTICAL DEVICE

The above research confirmed the mode of operation of the electrostatic design and led to a second set of experiments with rectangular transducers whose lateral dimensions could be changed easily and which could be constructed simply by attracting two aluminum membranes to a central paper layer, using the applied dc bias signal. These acoustic sources were thus extremely simple to fabricate using inexpensive materials. The aim of this second set of measurements was to determine the response of these devices and to measure their directivity.

Figure 13 is a photograph showing such a device, together with the measurement microphone. This could be scanned in a horizontal plane using a purpose-built scanning stage, controlled by LabVIEW™ software. The rest of the instrumentation was similar to that shown earlier in Fig. 10. All scans were initiated 3 mm (±0.5 mm) from the front face of the transducer and were performed using a 4-mm step size in both the x and z directions. The recorded field scans were analyzed by extracting the peak-to-peak amplitude and positional data from the measured data. This information was then plotted to create an image of the radiated pressure fields from the transducer under test.
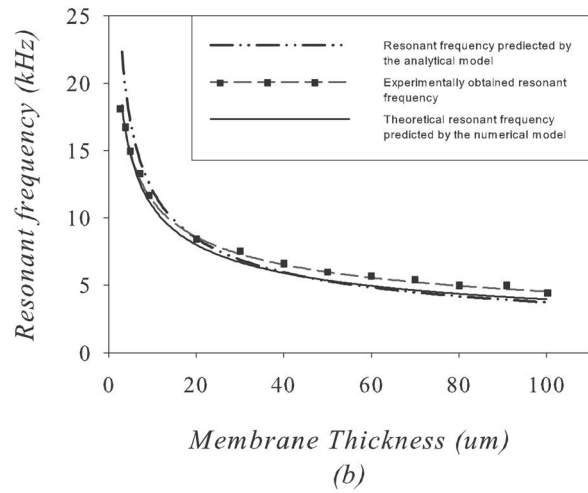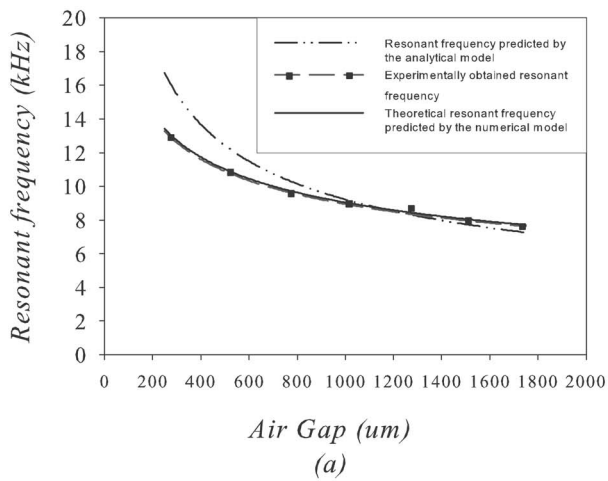
FIG. 12. Predictions of both mathematical models and experiment for the value of the first resonant frequency (f_1), as a function of changes in (a) the air gap for a constant membrane thickness and (b) the membrane thickness for a constant air gap.

The microphone was now scanned spatially in the horizontal (x-z) plane. Figure 14 shows the plots obtained from the same device when driven by a toneburst at (a) 10 kHz
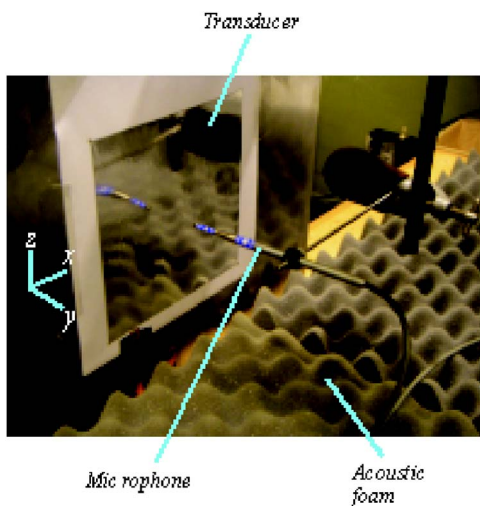
FIG. 13. Photograph of a simple rectangular electrostatic device, showing also the Brüel and Kjær 4138-A microphone mounted onto the scanning stage.
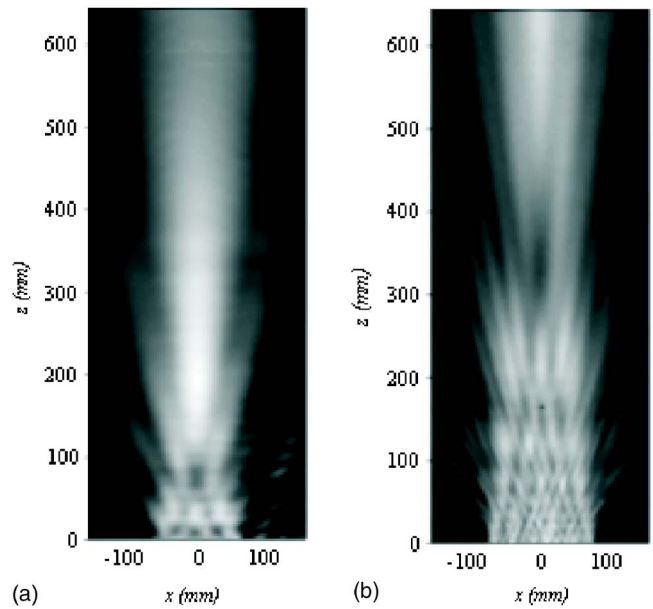
FIG. 14. Experimental radiated field from a 150-mm-square electrostatic source, driven by a toneburst at (a) 10 kHz and (b) 40 kHz.

and (b) 40 kHz. The data is displayed as a gray scale image in which the lighter shades represent higher received acoustic amplitudes. Note the greater number of side lobes and greater directivity at the higher frequency, as expected.

It is interesting to compare these results to those expected theoretically from a plane rectangular piston. This can be achieved using several methods.[19–21] In this work, the Rayleigh integral and convolution approach to predict the theoretical pressure fields from a signal source has been used to produce a numeric model for circular and rectangular sources, based upon the solution defined by Stepanishen[21] and San Emeterio and Ullate,[22,23] respectively. In this procedure the source is considered to act as a planar piston within an infinite baffle. Huygens' principle is then used, whereby the surface of the piston is imagined as being comprised of an infinite number of point sources that each produce a hemispherical pressure wave. In this way, the pressure field can be modeled using a surface integral over the area of the source. Figures 15(a) and 15(b) depict the resulting pressure amplitude for a 150-mm-square transducer, driven with 10- and 40-kHz toneburst signals. These signals were defined by the variable K in the formula used by San Emeterio and Ullate[23] to define a toneburst shape, which can be written as

$$\nu(t) = t^3 e^{-Kft} \cos(2\pi ft). \tag{18}$$

Here, f is the center frequency of oscillation and K is a constant that defines the bandwidth of the resultant waveform. A value of $K=0.4$ was used for both experimental and theoretical studies.

A comparison of Figs. 14 and 15 demonstrates good qualitative agreement between experimental results and the predictions for a plane piston. This is demonstrated further in Fig. 16, where the on-axis spatial variations in pressure amplitude at 10 kHz are plotted, both experimentally and as predicted by theory. Again, good agreement is evident.
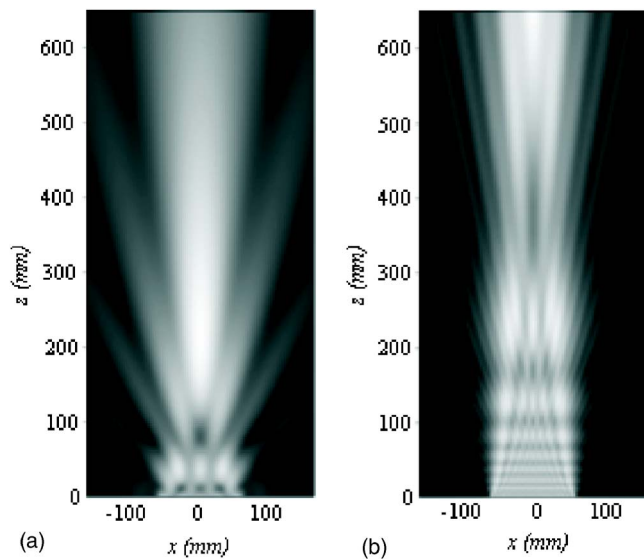
FIG. 15. Theoretical radiated fields for plane piston radiators of the dimensions and frequencies shown in Fig. 14.

V. DISCUSSION AND CONCLUSIONS

A new design of electrostatic transducer has been presented, which uses three solid layers separated by air gaps to produce acoustic signals over a wide frequency range. The basis of this design is that it is both thin and flexible and is hence different from those employed in both traditional electrostatic devices and capacitive devices used in the ultrasonic range.

An initial theoretical analysis showed that the use of three layers allowed more than one resonance to occur, because the central layer could move to create an asymmetric resonance. Experimental studies using a prototype device, in which the layer spacing and thickness could be easily adjusted, confirmed the predictions of theory from both numerical and analytical approaches. A feature of the design is thus that the bandwidth can be extended. In this first study, we have used symmetrical designs; however, it is possible to use different materials and thicknesses for each of the layers, to extend the frequency response and to make the response more uniform in amplitude over the required range. This is a feature of present work.
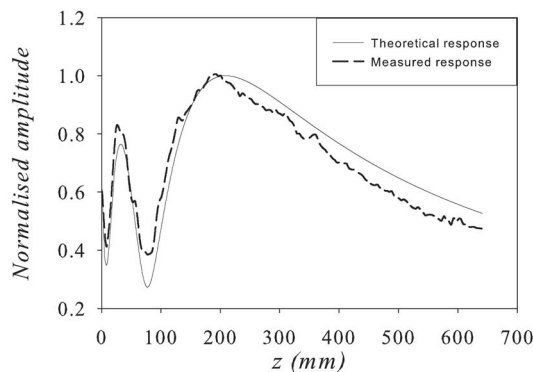
FIG. 16. Comparison of the measured and theoretical radiated pressure field along the z axis at $x=0$ for a 150–mm-square transducer when excited by a 10-kHz toneburst signal.

The experiments also appeared to confirm that the devices operated as plane pistons. This is to be expected if the force acting across their radiating area is reasonably uniform and that lateral (or "plate") resonances of each layer are absent. The use of thin layer materials and air gaps, coupled with wide lateral dimensions, discourages this behavior and encourage one-dimensional motion. The additional lack of clamping at the device edges also helps in this regard.

The electrostatic sources presented in this paper thus represent an interesting addition to the literature in that they behave like plane pistons and yet are simple to construct from easily available materials. Research is now being performed into their optimization in terms of bandwidth and sensitivity. One interesting aspect is that, being thin and flexible, they are easily contoured into different shapes. In addition, they can be constructed with large radiating surface areas, making them highly directional. These and other interesting aspects of their design are being actively pursued by the authors.

[1]E. C. Wente, "Absolute measurement of sound intensity," Phys. Rev. **10**, 39–63 (1917).
[2]P. E. Edelman and E. Banning, various US patents: "Acoustic condenser" (US Patent No. 1,759,809), "Loud-speaker condenser" (US Patent No. 1,776,112), "Condenser reproducer" (US Patent No. 1,759,810), "Push-pull condenser speaker" (US Patent No. 1,759,811), "Loud acoustic condenser" (US Patent No. 1,767,656), "Capacity reproducer" (US Patent No. 1,767,657), all published in 1930.
[3]A. A. Janszen, "An electrostatic loudspeaker development," J. Audio Eng. Soc. **3**, 87–90 (1955).
[4]E. W. Kellogg, General Electric. (1934). "Production of Sound," (US Patent No. 1983377).
[5]P. J. Walker, "New developments in electrostatic loudspeakers," J. Audio Eng. Soc. **28**, 795–799 (1980).
[6]J. D. Fox and B. T. Khuri-Yakub, "High frequency wave measurements in air," *Proc. 1983 IEEE Ultrasonics Symposium*, New York, 1983, 581–592.
[7]M. C. Bhardwaj, I. Neeson, and L. Neeson, "Ultrasonic analysis of plastics, rubbers and composites by non-contact analyzer—the NCA 1000," J. Aircraft Eng. Aerospace Technol. **71**, 39–41 (1999).
[8]G. Hayward and A. Gachagan, "An evaluation of 1–3 connectivity composite transducers for air-coupled ultrasonic alications," J. Acoust. Soc. Am. **99**, 2148–2157 (1996).
[9]W. Kuhl, G. R. Schröder, and F. K. Schröder, "Condenser transmitters and microphones with solid dielectric for airborne ultrasonics," Acustica **4**(5), 519–532 (1954).
[10]P. Mattila and J. Hietanen, "Bandwidth control of an electrostatic ultrasonic transducer," Sens. Actuators, A **45**, 203–208 (1994).
[11]M. J. Anderson, J. A. Hill, C. M. Fortunko, N. S. Dogan, and R. D. Moore, "Broadband electrostatic transducer: modeling and experiments," J. Acoust. Soc. Am. **97**, 262–271 (1995).
[12]E. I. Madaras, W. P. Winfree, B. T. Smith, and J. H. Heyman, "Detection of bondline delaminations in multilayer structures with lossy components," *Proc.1987 IEEE Ultrasonics Symposium*, Denver, Colorado, 1987, pp. 1047–1052.
[13]D. A. Hutchins, L. F. Bresse, and D. R. Billson, "Resonance studies of bonded aluminum joints," Nondestr. Test. Eval. **10**, 149–165 (1993).
[14]D. W. Schindel, "Air-coupled ultrasonic measurements of adhesively bonded mulit-layer structures," Ultrasonics **37**, 185–200 (1999).
[15]L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980).
[16]L. F. Bresse and D. A. Hutchins, "Elastic constant estimation using generation by pulsed lasers," J. Acoust. Soc. Am. **84**, 1751–1757 (1988).
[17]W. C. Young, *Roark's Formulas for Stress and Strain*, 6th ed. (McGraw-Hill, New York, 1989).
[18]B. H. Tongue, *Principles of Vibration* (Oxford U. P., Oxford, 1996).
[19]D. A. Hutchins and G. Hayward, "Radiated fields of ultrasonic transducers," in *Physical Acoustics, Vol. XIX*, edited by R. N. Thurston and A. D. Pierce (Academic, London, 1990).
[20]G. R. Harris, "Review of transient field theory for a baffled planar piston,"

Medley *et al.*: Properties of a thin electrostatic transducer

J. Acoust. Soc. Am. **70**, 10–20 (1981).

[21]P. R. Stepanishen, "Transient radiation of pistons in an infinite baffle," J. Acoust. Soc. Am. **49**, 1629–1638 (1971).

[22]J. L. San Emeterio and L. G. Ullate, "Diffraction impulse response of rectangular transducers," J. Acoust. Soc. Am. **92**, 651–662 (1992).

[23]J. L. San Emeterio and L. G. Ullate, "Diffraction impulse-response of rectangular transducers," J. Acoust. Soc. Am. **92**, 651–662 (1992).

The acoustic center of laboratory standard microphones

Salvador Barrera-Figueroa[a) and Knud Rasmussen[b)
Danish Primary Laboratory for Acoustics, Danish Institute for Fundamental Metrology,
Matematiktorvet B307, DK-2800 Lyngby, Denmark

Finn Jacobsen[c)
Acoustic Technology, Ørsted-DTU Technical University of Denmark, Ørsteds Plads B352,
DK-2800 Lyngby, Denmark

An experimental procedure is described for obtaining the effective acoustic distance between pairs of microphones coupled by a free field, leading to the determination of the position of the acoustic center of the microphones. The procedure, which is based on measuring the modulus of the electrical transfer impedance, has been applied to a large number of microphones. In all cases effects due to reflections from the walls of the anechoic chamber and the interference between the microphones have been removed using a time-selective technique. The procedure of determining the position of the acoustic center from the inverse distance law is analyzed. Experimental values of the acoustic center of laboratory standard microphones are presented, and numerical results obtained using the boundary element method supplement the experimental data. Estimated uncertainties are also presented. The results reported confirm values previously defined in an international standard and extend the frequency range. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2345830]

I. INTRODUCTION

The concept of acoustic center has been widely used in the development and practical realization of free-field reciprocity calibration of microphones.[1–3] The acoustic center of a microphone is defined as follows: "For a sound emitting transducer, for a sinusoidal signal of given frequency and for a specified direction and distance, the point from which the approximately spherical wavefronts, as observed in a small region around the observation point, appear to diverge."[4] In the case of a reciprocal transducer such as the condenser microphone the acoustic center is the same whether the microphone is used as transmitter or receiver. This equivalence is explained by the fact that the diffracted field is equivalent to the radiated field when the radiating/diffracting object is acting in either condition.[5]

The accuracy of the position of the acoustic center as a function of the frequency has a significant influence on the accuracy of the estimated free-field sensitivity. However, only a few attempts to determine this quantity have been found in the literature. Most results have been determined from deviations of the amplitude of the sound pressure from the inverse-distance law, but there have also been a few examples of methods of determining the acoustic center from phase measurements. Jacobsen *et al.* summarized the literature and showed that the various procedures for determining the position of the acoustic center give in general different results for most sound sources, including condenser microphones.[6]

The position of the acoustic center of the microphones can be obtained from measurements of the acoustic transfer impedance between pairs of microphones. The measurement of the transfer function between the microphones is carried out in an anechoic room. However, any real anechoic room is less than perfect, and small reflections from the walls and interference between the microphones will contaminate the transfer function making the estimate of the acoustic center less accurate. The nature of these unwanted effects has been analyzed using statistical methods,[7] and using time-selective procedures.[8] Although time-selective methods can eliminate the unwanted contaminations, application of such methods may affect the accuracy of the estimated acoustic center. In this paper the time-selective technique described in Ref. 8 has been used for removing the unwanted reflections and interference between the microphones. The effect of the time-selective technique on the electrical transfer impedances used for determining the acoustic center is analyzed.

Prediction of the position of the acoustic center requires a precise knowledge of the sound field generated by the microphone when this acts either as a sound source or as a receiver. Thus a precise knowledge of the displacement of the diaphragm of the microphone and its geometry is needed. Some attempts to determine the sound field when a microphone is introduced in the propagation path of a plane wave have been described in the literature.[9–13] However, these studies have been focused on obtaining the free-field correction of the microphone, except Ref. 11, which provided values of the position of the acoustic center of laboratory standard microphones of types LS1 and LS2 determined numerically using the boundary element method (BEM).

The purpose of this paper is to present and examine experimental values of the position of the acoustic center of

---
a)Electronic mail: sbf@dfm.dtu.dk
b)Electronic mail: kr@oersted.dtu.dk
c)Electronic mail: fja@oersted.dtu.dk

laboratory standard microphones based on deviations from the inverse-distance law. The experimental results are supplemented with numerical results obtained using the BEM.

II. DETERMINING THE ACOUSTIC CENTER FROM THE INVERSE-DISTANCE LAW

The discussion may be simplified if the microphones are regarded as axisymmetric sources observed from positions on the axis of symmetry. Under this assumption, the acoustic center must be somewhere on the axis. If the amplitude of the sound pressure is plotted as a function of the distance, a straight line can be fitted over the region of concern. Thus, the position of the acoustic center, $x(k,r)$, can be determined using the following expression:[6]

$$x(k,r) = r + |p(r)| / \frac{\partial |p(r)|}{\partial r}, \qquad (1)$$

where k is the wave number, r is the axial distance from the diaphragm of the microphone, $p(r)$ is the sound pressure as a function of distance, and the rate of change, $\partial |p(r)| / \partial r$, must be estimated by any available means, for example by using least squares fitting.

The rate of change can be determined if measurements of the electrical transfer impedance are made at several distances. This procedure can be used when the characteristics of the air in the environment where the measurements are carried out can be calculated accurately using well defined standard procedures. The sound pressure generated by a microphone used as a sound source decreases as a function of the distance from the acoustic center, r_t, following the inverse-distance law,

$$p(r) = j \frac{\rho f}{2r_t} M_f \cdot i \cdot e^{-\gamma r_t}. \qquad (2)$$

In Eq. (2) ρ is the density of air, M_f is the free-field sensitivity, f is the frequency, i is the current feeding the microphone, and γ is the complex propagation coefficient. This coefficient is defined as $\gamma = jk + \alpha$, where α is the air absorption. Thus, if an ideal receiver that does not disturb the propagating wave is located at a given distance, the output voltage will be inversely proportional to the true acoustic distance.

A realistic implementation of the above procedure must take account of the air absorption and the fact that the receiver will affect the sound field and will have its own acoustic center. In a reciprocity arrangement, the electrical transfer impedance between the microphones contains information about the position of the acoustic center of each microphone; therefore, it can be used for determining the sum of the acoustic centers of the two microphones. The electrical transfer impedance is defined as the ratio of the open-circuit voltage at the terminals of the receiver microphone, u_1, to the current through the terminals of the transmitter microphone, i_2,[4]

$$Z_{e,12} \equiv \frac{u_1}{i_2} = j \frac{\rho f}{2r_t} \cdot M_{f,1} \cdot M_{f,2} \cdot e^{-\gamma \cdot r_t}. \qquad (3)$$

In what follows the position of the acoustic center is defined as positive when the center is placed in front of the microphone diaphragm. Thus, the distance r_t in Eq. (3) is the difference obtained by subtracting the sum of the acoustic centers of the microphones, x_{1+2}, from the physical distance between the diaphragms of the microphones, r. Since only the modulus of the transfer impedance is of concern, it can be assumed that the parameters are independent of the distance. Substituting the true distance by its two components and rearranging terms gives

$$\frac{1}{|Z_{e,12}| e^{\alpha \cdot r_t}} \approx \frac{1}{A}(r - x_{1+2}) \approx mr + b. \qquad (4)$$

Thus, the sum of the acoustic centers, x_{1+2}, can be obtained by dividing the independent term, b, by the slope, m. The slope corresponds to the rate of change defined by Eq. (1). Note that the procedure is based on the following assumptions: (i) The observation distance is long compared to the sum of the acoustic centers. Under typical measurement conditions, the distances range from 250 mm to 500 mm for LS1 microphones, and from 160 mm to 350 mm for LS2 microphones. According to values given in the standard, the acoustic centers are about 9 mm at low frequencies for LS1 microphones, and about 5 mm for LS2 microphones.[4] (ii) The physical properties of the air inside the anechoic chamber can be estimated accurately using standard methods.[4,14] (iii) In the range of distances where the measurements are made, the position of the acoustic center varies very little.[15] The acoustic center determined by means of the above procedure is the average acoustic center at the average distance.

Typically, in a reciprocity measurement, the transfer impedance is measured using three microphones successively coupled in a free field. This yields three sums of acoustic centers of the three microphones. Thus, the position of the acoustic center of each microphone can be obtained by solving the resulting system of equations. The acoustic center of microphone No. 1 is

$$x_1 = \frac{b_{12}/m_{12} + b_{13}/m_{13} - b_{23}/m_{23}}{2}, \qquad (5)$$

where $m_{i,j}$ and $b_{i,j}$ are the slope and the independent term determined for the combination between microphones i and j, respectively. Similar expressions can be obtained for the other microphones.

III. NUMERICAL MODELING

The procedures described above have been applied onto experimental data. However, since there are discrepancies between reported results, some source of validation is needed. No analytical solution of the scattering/radiation problem of the microphone exists, but several numerical formulations have been developed for solving the problem. Any of these formulations can be used for comparison with the experimental results. In this case, an axisymmetric formula-
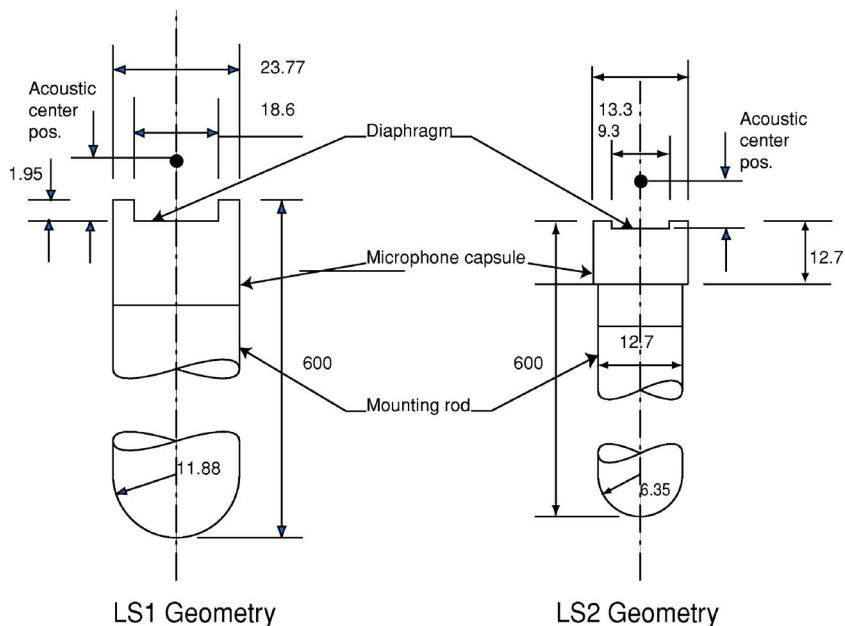
FIG. 1. Geometry of LS1 and LS2 microphones used in the BEM calculations. Dimensions in mm.

tion of the BEM (Ref. 11) is used for obtaining a numerical estimate of the position of the acoustic center.

This formulation was used by Juhl for determining the position of the acoustic center of laboratory standard microphones. Juhl's purpose was not to analyze the microphone problem in depth, but to validate his formulation with experimental data, and his low frequency approximation was based on the assumption that the movement of the diaphragm follows the first mode of the Bessel function of zero order and can be approximated by a parabolic function at frequencies below the resonance of the microphone. Under this assumption he calculated the position of the acoustic center up to 10 kHz. The assumption of the parabolic movement breaks down at higher frequencies where experimental results show significant deviations from the predicted values.

A modal sum of Bessel functions has been used extensively for describing the movement of the diaphragm.[9,10,12,13,16–18] This has yielded some fairly good approximations in calculations of the free-field correction, but no calculation of the acoustic center based on these results has been made.

In this investigation it has been assumed that the displacement of a diaphragm of radius a under uniform pressure conditions is

$$\eta(r_D) = \left[\frac{J_0(k_D r_D)}{J_0(k_D a)} - 1 \right] \eta_0, \qquad (6)$$

where J_0 is the Bessel function of zero order, η_0 is a constant that defines the amplitude of the movement, k_D is the wave number of the diaphragm, and r_D is the radial coordinate. A constant (pistonlike) displacement distribution has also been examined.

The geometry used in the simulations is shown in Fig. 1. The semi-infinite rod was approximated by a length of 60 cm. This will introduce a small disturbance in the results because of reflections from the back of the rod, but it is expected that the amplitude of this disturbance is negligible. The frequency range used in the calculations was 1 kHz–32 kHz for LS1 microphones and 4 kHz to 50 kHz for LS2 microphones. The size of the smallest element in the axisymmetric mesh was 2.5 mm and 1.5 mm for LS1 and LS2 microphones, respectively. Thus, there were at least four elements per wavelength at the highest frequency. It is well known that BEM solutions for exterior problems can be contaminated by spurious results associated with fictitious eigenfrequencies in the internal domain of the geometry.[19] This is known as the nonuniqueness problem. In the case at hand this problem can occur above about 10 kHz. The nonuniqueness problem has been avoided by adding a random CHIEF point as described in Ref. 19 and further checked by calculating the condition numbers of the BEM matrices[20] and by repeating calculations with small frequency shifts.

IV. EXPERIMENTAL RESULTS

The positions of the acoustic centers of a number of condenser microphones of types LS1 and LS2 have been determined using the linear regression procedure outlined above. Equation (4) gives the sum of the acoustic centers of the two microphones of the measured pair. The individual values are determined using Eq. (5).

The measurements were made using the experimental apparatus described in Ref. 21. The measurement system consists of a homemade reciprocity apparatus which is connected to a Brüel & Kjær PULSE (B&K) analyzer; the receiver and transmitter microphone are connected to the reciprocity apparatus. The environmental conditions are measured and recorded during the measurement of the electrical transfer impedances; temperature and relative humidity are kept at $23° \pm 0.5°$ and $50\% \pm 10\%$, respectively, and the static pressure is 100 kPa±2 kPa. A simplified diagram of the measurement system is shown in Fig. 2. The determination of the acoustic center involves the measurement of the electrical transfer impedance of the three pairs of microphones at four different distances (24, 32, 40, and 48 cm for LS1 microphones, and 16, 20, 24, and 30 cm for LS2 micro-
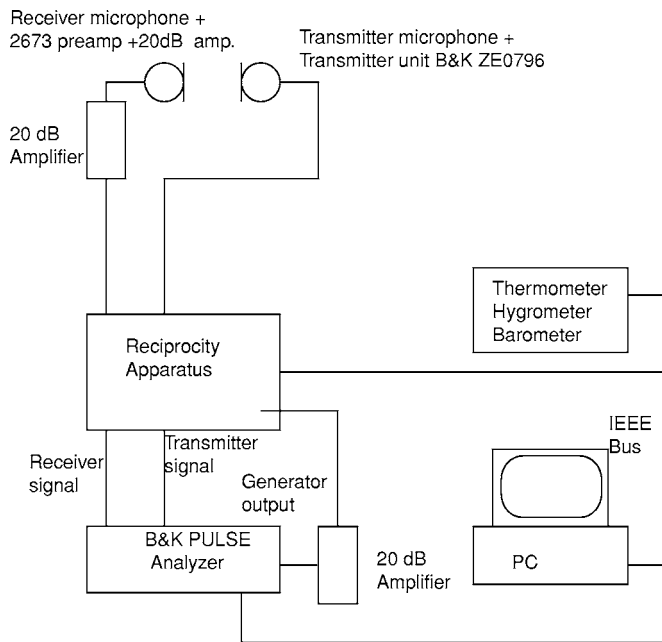
FIG. 2. Instrumentation used in the experimental determination of the acoustic center of Laboratory Standard Microphones.

FIG. 3. Comparison between "cleaned" (solid line) and "raw" (dashed-dotted line) estimates of the position of the acoustic center for LS1 and LS2 microphones. (a) LS1 microphones (B&K 4160); (b) LS2 microphones (B&K 4180).

phones), at discrete frequencies (from 1 kHz to 30 kHz for LS1 microphones, and from 3 kHz to 50 kHz for LS2 microphones) using stepped sine excitation with a frequency step of 120 Hz. The open-circuit voltage of the receiver microphone and the voltage on the terminals of the reference impedance connected in series with the transmitter microphone are measured using the B&K PULSE analyzer.

After some manipulations carried out onto the measured frequency response (described in Ref. 8), its associated impulse response is obtained by applying an inverse Fourier transform. The frequency step of 120 Hz is fine enough to obtain an impulse response that contains the primary reflections from the walls of the anechoic room and the interference between the microphones. These reflections are eliminated by applying a Tukey time-selective window with a duration of 1.8 ms for LS1 microphones and 1 ms for LS2 microphones. A direct Fourier transform is then applied to the time-windowed impulse response.

Once the influence of extraneous reflections from the walls and second-order interference between microphones have been removed from the measured electrical transfer impedance, the individual acoustic centers can be determined.

A. Effect of the time-selective procedure

Figure 3 shows a comparison between "cleaned" and "raw" estimates of the positions of acoustic centers of LS1 and LS2 microphones. Although the combined effect of cross-talk, interference between the microphones and reflections from the walls can be identified from the periodicity observed in the frequency domain, the identification of such effects is carried out more precisely in the time domain. A more detailed analysis can be found in Refs. 22 and 21. It is also clear that the time-selective procedure effectively removes much of the perturbations from the electrical transfer
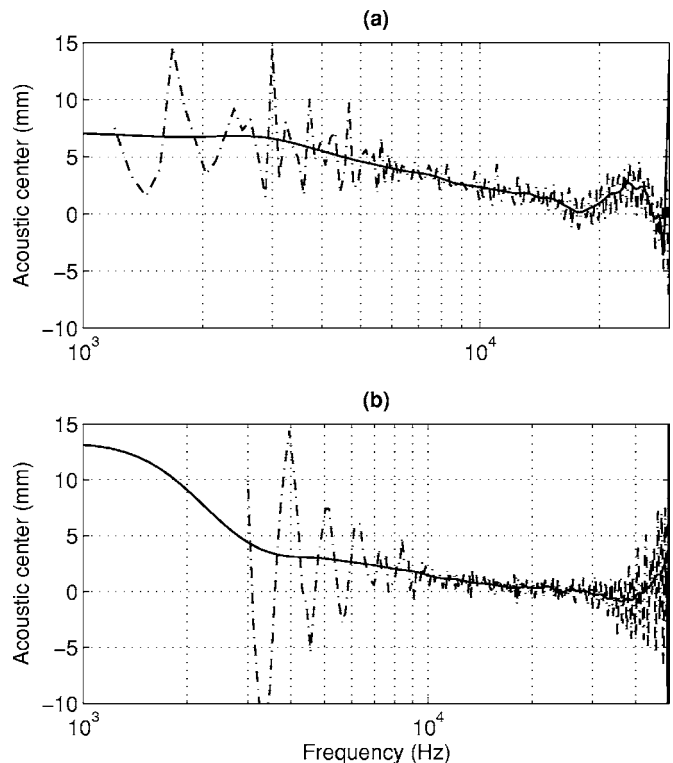
impedance and therefore from the estimate of the position of the acoustic center. The ripple observed at the extremes of the frequency range is caused by the time-selective window. Although this has a degrading influence on the estimated acoustic center, it seems to be preferable to the nonpredictable effect of the contaminations present in the raw estimate.

B. Average position of the acoustic center

Figure 4 shows the average acoustic center of LS1 and LS2 microphones (of 12 and 21 microphones, respectively) together with the standardized values.[4] It can be seen that the position of the acoustic center of LS1 microphones more or less follows the standardized values of the acoustic center in most of the frequency range, but above 15 kHz and up to 20 kHz it seems to diverge. The acoustic center reaches a maximum about 26 kHz, and then decreases again. This behavior for LS1 microphones was also observed in one laboratory's measurements in the international comparison,[23] although measurements were made only up to 25 kHz. The same tendency is observed for the LS2 microphones around 47 kHz. A similar behavior at high frequencies was registered for LS2 microphones in Ref. 7.

The frequency where the maximum occurs coincides with the frequency of the second radial resonance of the diaphragm. This suggests that the phenomenon may be related to the actual displacement of the diaphragm. This behavior will be compared with numerical BEM results.
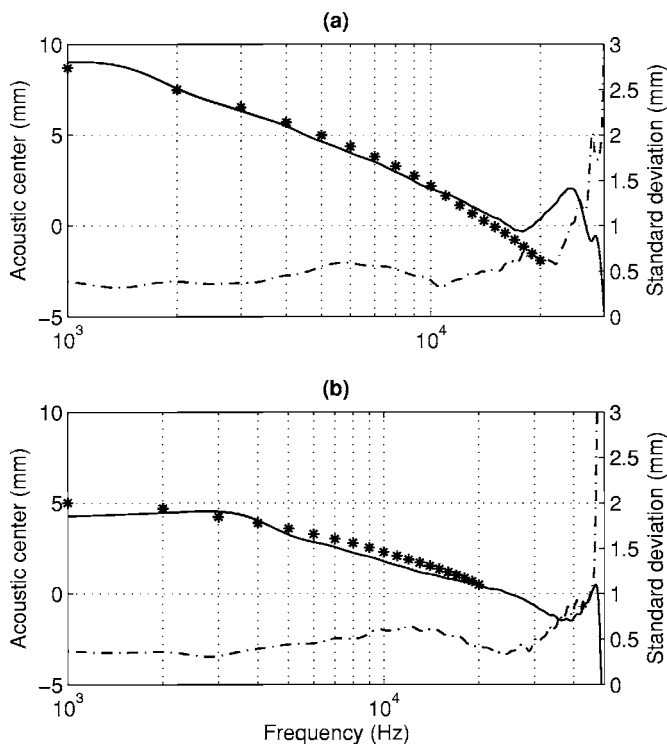
FIG. 4. Average position (solid line) and standard deviation (dashed-dotted line) of the acoustic center of a number of (a) LS1, and (b) LS2 microphones obtained experimentally. For comparison purposes, the standardized values are also included (stars).

FIG. 5. Acoustic center of (a) LS1 and (b) LS2 microphones determined experimentally and numerically using BEM. Solid line, experimental results; dotted line, uniform distribution; dashed line, Bessel distribution.

V. NUMERICAL RESULTS

Figure 5(a) shows the results of the numerical calculations of the position of the acoustic center of LS1 microphones using the axisymmetric BEM formulation. It is interesting to note that the high frequency behavior of the calculated acoustic center resembles the experimental results. At low frequencies the assumed shapes of the displacement yield an acoustic center that follows the experimental estimate, although the uniform displacement distribution gives a slightly smaller value. It can also be seen that the calculated acoustic centers have a maximum at about the same frequency, which is the second radial resonance of the diaphragm. It is interesting to observe that in the frequency range where the peak is observed, it is the uniform distribution that shows the best agreement with the experimental values. This may be caused by the fact that at such high frequencies, the effect of the air film between the diaphragm and the back plate of the microphone begins to dampen the amplitude of the movement of the diaphragm, perhaps heavily, in the middle of the diaphragm, thus flattening the displacement distribution and making it look more like the uniform distribution. A similar behavior is observed in the case of LS2 microphones shown in Fig. 5(b). It should be interesting to have measured values of the actual velocity distribution of the diaphragm.

VI. DISCUSSION

An aspect of interest is the effect of the distance range in which the experimental acoustic centers are determined. This dependence has been demonstrated for some sources,[6] and it

would not be surprising to find a similar behavior in the case of the microphones.[11] Figure 6(a) shows the position of the acoustic center of LSI microphones calculated numerically at different average distances: 240 mm, 480 mm, and 1000 mm. It can be seen that there is a systematic influence of the distance on the position of the acoustic center. It should converge to a value at the largest distances, a far-field acoustic center. At low frequencies (below 1 kHz) the acoustic center converges slowly while above this frequency the convergence is faster.

Although it is very difficult to make reliable measurements of the electrical transfer impedance at large distances between the microphones because the signal to noise ratio is very poor, an experiment was carried out in order to estimate the effect of the distance on the estimated acoustic center. Figure 6(b) shows the experimental acoustic center at average distances of 260 mm and 380 mm. It can be seen that the acoustic center moves away from the microphone when the

TABLE I. Coefficients of the polynomial approximation of the acoustic center.

LS1	Coeff.	LS2
10.8	a_0	4.6
−2.32	a_1	−1.13
0.317	a_2	−0.052
−0.0268	a_3	0.0055
0.0011	a_4	−0.000224
−1.64e−5	a_5	3.99e−6
...	a_6	−2.573e−8
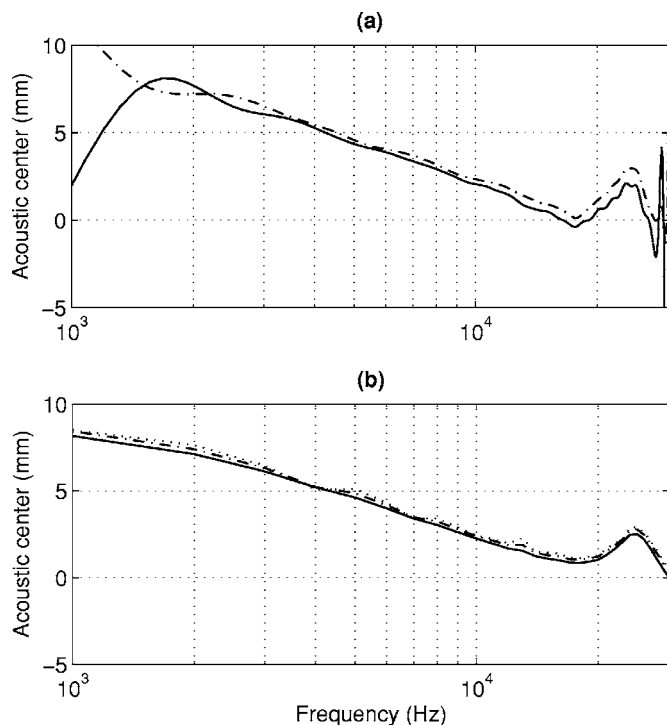
FIG. 6. Position of acoustic center of LS1 microphones obtained at different distances. (a) BEM results: Solid line, 240 mm; dashed-dotted line, 480 mm; and dotted line, 1000 mm. (b) Experimental values: Solid line, 260 mm; dashed-dotted line, 380 mm.

average distance is increased. It can be expected that the position of the acoustic center converges to a far-field value as the average distance increases, as observed in the numerical calculations.

The convergence to a far-field acoustic center leads to a discussion about which acoustic center should be used in free-field reciprocity calibration of microphones. An important consequence of the use of the acoustic center determined from modulus measurements is that the resulting free-field sensitivity becomes independent of the calibration distance.

Hence, if small changes in the acoustic center are introduced, this independence is degraded, and this will lead to changes in the free-field sensitivity. A change of 1 mm in the position of the acoustic center will lead to a change of 0.03 dB in sensitivity when the measurements are carried out at an average distance of 300 mm. One possibility would be to use the acoustic center determined experimentally, and to obtain a sensitivity that is valid for the distance range at which the measurements are made. Afterward, a correction could be made in order to obtain a far-field microphone sensitivity.

Because of the poor signal to noise ratio, it is very difficult to determine the position of the acoustic center experimentally at low frequencies and long distances. Additionally, the use of time-windowing will degrade the accuracy of the estimate of the acoustic center position at low frequencies. However, values of the acoustic center in the whole frequency range may be needed. One possible solution is to determine a polynomial approximation of the acoustic center that is obtained from experimental values at mid and high frequencies and from numerical values in the lowest frequency range.

If an accurate prediction of the acoustic centers at high frequencies using BEM is wanted, it is crucial to use a displacement distribution that represents the actual movement. Recent work suggests that a LASER vibrometer may be used to measure the displacement distribution.[24] This shape can subsequently be used to determine the acoustic center at frequencies where the displacement of the membrane is no longer described using simple functions.

Finally, as the position of the acoustic center of an individual microphone is not expected to change unless there is a significant change of the acoustic impedance of the microphone, any convenient choice of a polynomial approximation can be used advantageously instead of determining the experimental values every time. Coefficients of polynomial approximations of the experimental results are shown in Table I. These coefficients are used in the following polynomial:

TABLE II. Sources of uncertainty of the acoustic center for LS1 and LS2 microphones.

	LS1 microphones													
f (kHz)	1	2	3	4	5	10	15	20	25	26	27	28	29	30
Voltage ratio (dB), Type A	0.035	0.026	0.017	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.017	0.026	0.035
Cleaning technique (dB), Type B	0.146	0.026	0.009	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.009	0.026	0.146
Air attenuation (dB), Type B	0.000	0.000	0.000	0.001	0.001	0.005	0.010	0.015	0.018	0.019	0.019	0.019	0.019	0.020
Distance (mm), Type B	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Reproducibility (mm), Type A	0.39	0.41	0.33	0.43	0.53	0.37	0.44	0.63	0.76	0.71	0.84	1.7	1.2	1.8
Far field (mm), Type B	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

	LS2 microphones													
f (kHz)	1	2	3	4	5	10	20	30	40	45	46	47	48	49	50
Voltage ratio (dB), Type A	0.043	0.026	0.017	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.017	0.026	0.035
Cleaning technique (dB), Type B	0.206	0.026	0.017	0.013	0.007	0.001	0.001	0.001	0.001	0.001	0.007	0.013	0.017	0.026	0.206
Air attenuation (dB), Type B	0.0000	0.0000	0.0000	0.0001	0.0001	0.0004	0.0011	0.0014	0.0014	0.0015	0.0015	0.0015	0.0016	0.0016	0.0017
Distance (mm), Type B	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Reproducibility (mm), Type A	0.32	0.3	0.31	0.38	0.44	0.6	0.4	0.44	0.66	0.85	0.9	1.05	2.6	2.6	2.6
Far field (mm), Type B	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Barrera-Figueroa *et al.*: Acoustic center of microphones   2673

TABLE III. The position of the acoustic center and its uncertainty for LS1 microphones at selected frequencies.

f (kHz)	Ac. center (mm)	u (mm)	f (kHz)	Ac. center (mm)	u (mm)
1.0	9.01	5.6	16.0	0.14	0.6
2.0	7.56	1.5	17.0	−0.21	0.6
3.0	6.30	0.9	18.0	−0.28	0.7
4.0	5.43	0.6	19.0	−0.03	0.7
5.0	4.59	0.6	20.0	0.36	0.8
6.0	3.95	0.6	21.0	0.80	0.9
7.0	3.46	0.6	22.0	1.20	1.0
8.0	2.92	0.6	23.0	1.69	1.1
9.0	2.41	0.6	24.0	1.99	1.3
10.0	2.03	0.6	25.0	1.84	1.4
11.0	1.77	0.6	26.0	0.93	1.5
12.0	1.43	0.6	27.0	−0.33	1.7
13.0	1.03	0.6	28.0	−0.57	1.9
14.0	0.71	0.6	29.0	−1.07	2.5
15.0	0.44	0.6	30.0	−4.63	6.7

$$x(f) = a_0 + a_1 f + a_2 f^2 + \cdots + a_n f^n, \tag{7}$$

where f is in kHz.

VII. CONCLUSIONS

The problem of determining the position of the acoustic centers of laboratory standard microphones has been addressed, and the acoustic center of laboratory standard microphones, LS1 and LS2, has been determined from the modulus of the electrical transfer impedance measured at different distances between pairs of microphones.

TABLE IV. The position of the acoustic center and its uncertainty for LS2 microphones at selected frequencies.

f (kHz)	Ac. center (mm)	u (mm)	f (kHz)	Ac. center (mm)	u (mm)
1.0	4.26	5.71	26.0	−0.09	0.58
2.0	4.44	1.83	27.0	−0.18	0.58
3.0	4.56	0.73	28.0	−0.33	0.58
4.0	4.04	0.62	29.0	−0.50	0.58
5.0	3.25	0.60	30.0	−0.64	0.58
6.0	2.84	0.62	31.0	−0.81	0.59
7.0	2.58	0.64	32.0	−0.92	0.60
8.0	2.26	0.66	33.0	−1.12	0.61
9.0	2.04	0.68	34.0	−1.20	0.63
10.0	1.80	0.70	35.0	−1.29	0.65
11.0	1.58	0.70	36.0	−1.45	0.69
12.0	1.38	0.70	37.0	−1.46	0.71
13.0	1.18	0.70	38.0	−1.44	0.73
14.0	1.08	0.69	39.0	−1.52	0.75
15.0	0.96	0.68	40.0	−1.37	0.75
16.0	0.84	0.66	41.0	−1.21	0.78
17.0	0.77	0.64	42.0	−1.15	0.81
18.0	0.66	0.62	43.0	−0.68	0.84
19.0	0.60	0.61	44.0	−0.58	0.87
20.0	0.51	0.60	45.0	−0.32	0.95
21.0	0.35	0.60	46.0	0.20	1.10
22.0	0.27	0.59	47.0	0.52	1.20
23.0	0.24	0.59	48.0	−0.75	2.90
24.0	0.12	0.59	49.0	−6.23	3.20
25.0	−0.01	0.59	50.0	−17.57	10.20

The position of the acoustic center has also been estimated using an axisymmetrical BEM formulation assuming different velocity distributions of the diaphragm. At low frequencies and up to the resonance frequency of the microphone, the use of simple functions such as the Bessel function yields values that are in good agreement with the experimental values. At higher frequencies the agreement degrades but the numerical results nevertheless confirm the experimental estimates.

There is a good agreement between the standardized and experimental values of the acoustic center in the mid-frequency range. The discrepancies observed at frequencies up to 20 kHz for LS1 microphones, and the need of sensitivity estimates at higher frequencies may point to the need of updating the standard. The experimental results presented in this paper may be regarded as a contribution to this process.

APPENDIX: TABULATED VALUES OF THE POSITION OF THE ACOUSTIC CENTER AND ASSOCIATED UNCERTAINTIES

The uncertainty of the estimate of the position of the acoustic center can be determined by following the ISO/GUM guide for estimating the uncertainty of measurements.[25] Equation (5) is the mathematical model used in the estimation of the uncertainties. The estimated slope and independent term are determined from transfer impedances "cleaned" from the influence of reflections and corrected for the effect of the air absorption. A list of the possible contributors is given below.

1. **Electrical transfer impedance:** The modulus of the electric transfer impedance can be affected by many factors. The most conspicuous are the distance between microphones, the accuracy in the measurement of the voltage ratio, and the environmental conditions. The voltage ratio is measured using a steady state response (SSR) analyzer. This makes it possible to measure the frequency response with a given accuracy using a stepped sine excitation. The measurement is made using two channels in the same input range.
2. **Time-selective technique:** The time-selective technique removes disturbances from the frequency response very effectively, but it also introduces ripples in the extremes of the frequency response. The ripple is caused by the fact that the time-selective window truncates the impulse response at positions where it is not zero. Thus, the ripple is a function of the length and position of the time window.
3. **Microphone sensitivity:** The microphone free-field sensitivity may change slightly as a function of distance, but for the purpose of estimating the uncertainty of the position of the acoustic center of an individual microphone, it can be regarded as constant. However, in the uncertainty of the average position of the acoustic center of a number of microphones, deviations related to the geometry and acoustic impedance should be taken into account.
4. **Air absorption:** The electrical transfer impedance is corrected for the losses induced by air absorption. This

is a function of some basic constants and the environmental conditions and the *physical* distance between the diaphragms of the microphones rather than the distance between acoustic centers. This is due to the fact that the acoustic waves actually propagate between the two diaphragms. The difference between using the physical distance and the acoustic distance is negligible in the frequency range of interest, though.

5. **Least-squares fitting:** The slope and the intercept are estimated using the least squares technique. These estimates have an associated standard error.

These contributors can be added to the model based on Eq. (5). Further simplification can be achieved by assuming that the value and uncertainty of each intercept and slope is the same. Thus, the mathematical model is

$$x_1(f) = \frac{b/m}{2} + R + LS + FF. \tag{A1}$$

The least-squares fitting is made over the electrical transfer impedance measured at different distances. Each electrical transfer impedance is corrected for air absorption, thus $b, m = f\{Z_e, e^{\alpha r}\}$, R is the reproducibility, LS is the contribution from the least-squares fitting, FF is the contribution due to the difference between the far field acoustic center and the acoustic center estimated at short distances between the pair of microphones; this quantity can be estimated experimentally or numerically as described in Sec. VI.

The uncertainty of the "cleaned" electrical transfer impedance measured at each distance can be estimated using a model that includes the effect of the measurement procedure, the time-selective technique, the distance, and the air absorption.

The uncertainty of the total absorption, which depends on the environmental conditions and the distance between the diaphragms of the microphones, can be estimated on the basis of the calculation procedure described in the standard,[4] or the procedure described in a report that harmonized the calculation methodologies of the air properties used in the calibration of microphones.[14] The analytical expressions can be difficult to differentiate, and a simplified procedure based on finite differences[25] has been applied instead. The uncertainty of the temperature, relative humidity, and static pressure is estimated to be 0.5°, 5%, and 50 Pa, respectively.

The inclusion of the reproducibility may result in overestimation of the uncertainty because it may contain the effect of factors already considered in the other contributors. However, as the present model does not include all the possible contributors, it is preferable to have a conservative estimate of the uncertainty of the acoustic center.

The contribution of the least squares fitting can be obtained using the estimate of the standard error. However, in this exercise, the correlation coefficient is always better than 0.99. Therefore, the standard error can be considered negligible. Table II shows the sources of uncertainty. Once the contribution of each element is established, the standard uncertainty of the acoustic center has been estimated using the procedures described in the ISO/GUM.

Table III gives values of the position of the acoustic center and their uncertainty for LS1 microphones at selected frequencies. Table IV lists the corresponding values of the acoustic center and its uncertainty for LS2 microphones.

[1] W. R. MacLean, "Absolute measurement of sound without a primary standard," J. Acoust. Soc. Am. **12**, 140–146 (1940).

[2] W. Wathen-Dunn, "On the reciprocity free-field calibration of microphones," J. Acoust. Soc. Am. **21**, 542–546 (1949).

[3] I. Rudnick and M. N. Stein, "Reciprocity free-field calibration of microphones to 100 kc in air," J. Acoust. Soc. Am. **20**, 818–825 (1948).

[4] IEC International Standard 61094-3, Measurement microphone Part 3: Primary method for free-field calibration of laboratory standard microphones by the reciprocity technique (1995).

[5] F. M. Wiener, "On the relation between the sound fields radiated and diffracted by plane obstacles," J. Acoust. Soc. Am. **23**, 697–700 (1951).

[6] F. Jacobsen, S. Barrera-Figueroa, and K. Rasmussen, "A note on the concept of acoustic center," J. Acoust. Soc. Am. **115**, 1468–1473 (2004).

[7] R.-P. Wagner and V. Nedzelnitsky, "Determination of acoustic center correction values for type LS2aP microphones at normal incidence," J. Acoust. Soc. Am. **104**, 192–203 (1998).

[8] S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "A time-selective technique for free-field reciprocity calibration of condenser microphones," J. Acoust. Soc. Am. **114**, 1467–1476 (2003).

[9] E. Matsui, "Free-field correction for laboratory standard microphones mounted in a semi-infinite rod," J. Acoust. Soc. Am. **49**, 1475–1483 (1971).

[10] L. Bjørnø, O. Holst-Jensen, and S. Lolk, "Free-field corrections of condenser microphones," Acustica **33**, 166–173 (1975).

[11] P. M. Juhl, "A numerical investigation of standard condenser microphones," J. Sound Vib. **177**, 433–446 (1994).

[12] R. Barham, "Free-field reciprocity calibration of laboratory standard microphones," Ph.D. thesis, ISVR, University of Southampton, l995.

[13] X. Bao and Y. Kagawa, "A simulation of condenser microphones by boundary element approach," J. Sound Vib. **119**, 327–337 (1987).

[14] K. Rasmussen, "Calculation methods for the physical properties of air used in the calibration of microphones," Technical Report, No. PL-11a, Department of Acoustic Technology, Technical University of Denmark, 1996.

[15] K. Rasmussen, "Acoustic centers of condenser microphones," Technical Report, No. 5, The Acoustics Laboratory, Technical University of Denmark(1973).

[16] P. M. Juhl, "The boundary element method for sound field calculations," Ph.D. thesis, The Acoustics Laboratory, Technical University of Denmark (1993).

[17] A. J. Zuckerwar, "Theoretical response of condenser microphones," J. Acoust. Soc. Am. **64**, 1278–1285 (1978).

[18] Z. Skvor, *Vibrating Systems and their Equivalent Circuits* (Elsevier, New York, 1991).

[19] H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 45–58 (1968).

[20] P. M. Juhl, F. Jacobsen, V. Cutanda, and S. Quiros-Alpera, "On the non-uniqueness problem in a 2-D half-space BEM formulation," in *Proceedings of the Ninth International Congress on Sound and Vibration*, Orlando, FL (2002), cD-ROM.

[21] S. Barrera-Figueroa, "New methods for transducer calibration: Free-field reciprocity calibration of condenser microphones," Ph.D. thesis, Ørsted-DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark, 2003.

[22] S. Barrera-Figueroa, K. Rasmussen, F. Jacobsen, and L. Muniz, "On the interference between the two microphones in free-field reciprocity calibration," J. Acoust. Soc. Am. **116**, 2771–2778 (2004).

[23] K. Rasmussen and E. Sandermann-Olsen, "Intercomparison on free-field calibration of microphones," Technical Report, No. PL-07, The Acoustics Laboratory, Technical University of Denmark, 1993.

[24] G. Behler and M. Vorlander, "Reciprocal measurements on condenser microphones for quality control and absolute calibration," Acta. Acust. Acust. **90**, 152–160 (2004).

[25] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML, Guide to the expression of uncertainty in measurement (1995).

An investigation of surface-contacting sensors for the seismic detection of buried landmines

J. S. Martin and G. D. Larson
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

W. R. Scott, Jr.
School of Electrical and Computer Engineering, Georgia Institute of Technology,
Atlanta, Georgia 30332-0250

Techniques have been studied for the detection of buried landmines with acoustic/seismic interrogation signals. Much of this work has involved full wave-field imaging from local measurements of ground motion using noncontact sensors. These offer inherent safety for the system operator and accommodate the need to make measurements over rough ground surfaces. The system requirement is, however, only that a sensor does not intrude on the measurement rather than that it not contact the ground. An experimental investigation was conducted into the feasibility of an array of ground-contacting sensors for use in a seismic landmine-detection system that exploits full wave-field imaging. The main considerations in the design of the array sensor were safety, sensitivity, fidelity, reproducibility, and sensor-to-sensor interaction. A relatively simple and inexpensive sensor was demonstrated in an experimental simulation of a landmine-detection system. The sensor, which is suitable for inclusion in a large planar array that could be used for detection confirmation, exerts a safe normal force at the point of contact and enables detection performance comparable to that which could be achieved using noncontact techniques. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2345911]

I. INTRODUCTION

The use of acoustic/seismic signals to detect buried landmines has been studied for well over 30 years.[1] Some of the system concepts that have been explored involve remotely located arrays of ground-contacting seismic sensors and sources to detect mines using pulse-echo techniques.[2,3] Other system concepts involve local measurement of ground motion at locations, including some directly above buried mines using noncontact sensors.[4,5] For these systems, the motion is induced by either a remote seismic source or direct air-acoustic coupling at the measurement location. Local measurements offer the advantage that the signal scattered by a landmine is not diminished by geometrical spreading, absorption, or scattering from inhomogeneities in the medium before it can be measured. They also permit the entire surface wave-field to be captured so that data can be processed in the domains of both time and two-dimensional space in order to image the target (concepts that involve direct subsurface vibration measurement have also been proposed in the literature, although these have not been developed to a point of experimental testing[6]). Several noncontact sensor types have been found to be effective in systems based on local measurements. These include laser Doppler vibrometers (LDVs),[4,7,8] microwave radar-based vibrometers,[5,9] and ultrasonic vibrometers.[10,11]

The ground-contacting sensors studied here have been considered in the context of a previously investigated seismic landmine-detection technique.[5] This technique employs audio-frequency seismic surface waves. These waves are referred to as Rayleigh waves for the sake of brevity, although they are in fact the lowest order coupled P-SV mode in a vertically heterogeneous medium (i.e., stratified soil). Rayleigh waves interrogate only the near surface layers of soil because they decay exponentially into the medium.[12] These waves are excited by a remotely located source and the full wave-field is measured with a nonintrusive sensor that is scanned to form a synthetic two-dimensional array. The source that is most commonly used is an electrodynamic shaker that is coupled to the ground through a foot designed to preferentially excite Rayleigh waves. Landmine images are formed by post-processing the displacements of the ground's surface measured at many locations. The image processing, which involves spatial and temporal filtering, has been described in previous papers.[5,13]

In previous work, the receiving array has been synthesized using one or two sensors to scan thousands of measurement locations (typically 2500–5000).[5] The most significant drawback to this technique is the slow measurement speed that is forced by the synthesis of the receiving array. An obvious path to improve speed is by replacing the synthetic array with a physical array of sensors. Compatibility with this array requires individual sensors with low cost, small physical size, and low power consumption.

A physical array of sensors should yield time savings that are proportional to the number of elements in the array (*N*) because received signals are simultaneously acquired, and background noise integrated, at many measurement locations. If the array is moved incrementally, movement is *N* times less frequent for the scanning of a predefined area. In a realistic scenario, *N* should be an underestimate for the time

savings that can be realized with a physical array in comparison to a synthetic array of identical sensors. This is because the noise field is accurately sampled in time and space and can therefore be filtered more effectively. Nonseismic noise can be removed more effectively by spatial/temporal filters. Seismic noise that passes these filters should constitute a signal of opportunity.

There are many physical configurations that a field-operable landmine-detection system might take. Recent interest been focused on the development of confirmation systems, which are intended to quickly interrogate a prescribed area that has been identified as a possible landmine location. Here a large physical array of sensors may be used to span an entire region of interest, and the total measurement time should be only that which is required to build a sufficient signal to noise ratio for any individual sensor in the array.

II. ACCELEROMETER-BASED SENSORS

There are many technologies available for measuring surface motion with sensors that contact a surface of interest. These include accelerometers, velocimeters (geophones), seismometers, pressure sensors, and strain sensors. Commercially available accelerometers are an appealing alternative because of their low cost, variety, and ubiquity. Many of these have intrinsic noise densities that are superior to the radar-based sensor used in previous work over most of the frequency range of interest for landmine detection (50–1000 Hz). The radar-based sensor responds to displacement rather than acceleration, so there is a crossover point for signal to noise performance. Many small accelerometers have intrinsic noise density of $10-100$ μg/$\sqrt{\text{Hz}}$ in the frequency range of interest. This compares to 1 nm/$\sqrt{\text{Hz}}$, which was typical of both ultrasonic and radar-based sensors in bench-top testing. Therefore, the crossover for superior noise performance by an accelerometer is in the 50–150 Hz frequency range. Commercially available LDVs advertise intrinsic noise floors of 50 nm s^{-1}/$\sqrt{\text{Hz}}$. Compared to this, the crossover for superior noise performance by small accelerometers is in the 300–3000 Hz range. Soil is an inconsistent reflector for radar, ultrasound, and laser signals. Measurements using these modalities are usually limited by speckle noise rather than the intrinsic noise of noncontact sensors. Because of this, the crossover frequencies for superior noise performance by accelerometers, which are not susceptible to speckle noise, should be lower for practical purposes than they are for bench-top testing conditions.

An accelerometer is only functional if it is in intimate contact with a vibrating surface. The nature of that contact must satisfy conditions similar to noncontact sensors; the contact must not intrude on the vibration of the ground and it must reasonably couple surface motion to the accelerometer. Additionally, this coupling must be reproducible throughout an array of sensors and when an individual sensor is relocated.

There are three ways in which a ground-contacting sensor might intrude unacceptably on the soil surface in a landmine-detection system. First and most obviously, it could require a static biasing force that is sufficient to detonate a buried pressure-fused landmine. Anti-personnel (AP) mines impose a more stringent requirement than anti-tank (AT) mines in this regard. Typically pressure-fused AP mines detonate with contact forces of 20–160 N, although some AP mines may be detonated with applied forces as low as 10 N.[14] AT mines detonate with applied forces that are an order of magnitude higher than this. Second, a sensor may load either the ground or a buried mine to alter the surface motion. This is problematic if it degrades the contrast between landmines and background that can be generated by processing the collected data. Thus, some forms of ground loading may be more acceptable than others. Finally, a sensor may scatter the seismic interrogation signal and alter the signals that are received by other sensors in the array. These requirements generally dictate a sensor such as an accelerometer that is small, lightweight, and rigid.

The suitability of an accelerometer for use as a ground-contacting sensor is obvious. However, there are many suspension and packaging schemes that can be used to couple an accelerometer to the ground, and choosing among these is the more difficult question that was addressed in the design of the sensor. This problem was studied through an analysis involving simple lumped-element models for sensor response. The models were validated and several sensor designs were tested in a series of experiments, both in the laboratory and at a realistic field-test site. All of the sensor designs that were modeled and tested were based on commercially available accelerometers.

III. SENSOR DESIGN

The ground-contacting sensor design under consideration can be described as an accelerometer coupled to the ground's surface with a biasing force acting against a tail mass through a spring (as shown in the inset in Fig. 1). This configuration can be represented as a two degree-of-freedom (2-DOF) system.

The sensor head is modeled as a small mass (m_1) that is attached to the ground through a spring and damper (k_0 and η_0) that represent the reactive and resistive components of the ground's input impedance. Previous experiments have shown that the input reactance of the soil at various field sites is dominantly stiffness-like in the frequency range that is of interest for seismic landmine detection.[5] The stiffness of carefully compacted damp sand, which is a reasonable soil surrogate,[2,3,15–18] has been measured in the laboratory using an impedance head and found to be approximately 1 $\times 10^6$ N/m, with a circular contact area of 2 cm^2 that is typical of the prototype sensors. Although viscous damping was assumed, the frequency dependence of the ground-dominated sensor damping was not known and no physical mechanism was assumed for this process. The measured input resistance of the sand in the experimental model can be reasonably modeled in the frequency range of interest as a viscous damper with 76 N/(m/s) damping for a 2 cm^2 contact area. Laboratory measurements were best fit by a frequency dependence that related loss and displacement by a factor of $\omega^{0.84}$, which is reasonably close to the $\omega^{1.0}$ depen-
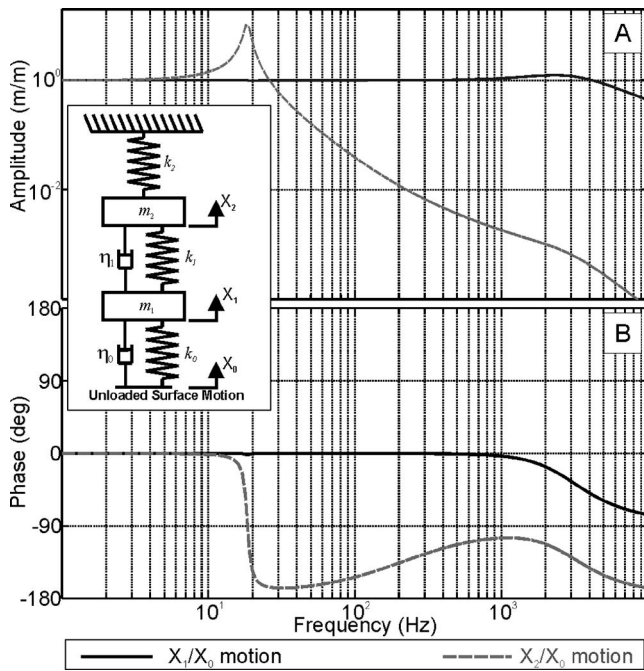
FIG. 1. Modeled response of a gravity-biased ground contacting sensor with a 156 g tail mass (A) amplitude and (B) phase. The model is shown in inset. Here the biasing spring $k_2=0$, foam collar $k_1=2.1\times10^3$ N/m, ground $k_0=1\times10^6$ N/m, collar damping $\eta_1=1.6$ N/(m/s), ground damping $\eta_0=76$ N/(m/s), head mass $m_1=5.6$ g, and tail mass $m_2=156$ g.

dence that would be expected for viscous damping. Field measurement data of a sufficient quality to perform a similar inversion for other soil types were not available. The tail mass is modeled as a lumped mass (m_2) that is coupled to the sensor head by means of a second spring (k_1). For some of the sensor designs that were examined, the tail mass was coupled to a backing structure through a third spring (k_2), which was used to provide an adjustable biasing force. In selecting parameters for the sensors, k_2 and m_2 were chosen to be sufficient to isolate the tail mass from motion of a complex supporting structure (an automated positioning system) in order to render the device as a whole insensitive to back-plane motion. An accelerometer is an absolute motion rather than a relative motion sensor, but back-plane motion may be coupled into the ground through the accelerometer and thereby affect accelerometer output. This isolation meant that, although the experiments involved a complex supporting structure, the stationary backing assumed in the model is a reasonable approximation. For all of the sensor prototypes, k_1 was selected to isolate m_2 and to be small in comparison to k_0. In order to minimize intrusion, m_1 was chosen to be as light as possible. The motions of the masses in this model are described by a system of equations as follows:

$$
\begin{bmatrix} k_0 + k_1 - m_1\omega^2 + i\omega\eta_0 + i\omega\eta_1 & -k_1 - i\omega\eta_1 \\ -k_1 - i\omega\eta_1 & k_1 + k_2 - m_2\omega^2 - i\omega\eta_1 \end{bmatrix}
$$
$$
\times \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} (k_0 + i\omega\eta_0)x_0 \\ 0 \end{bmatrix}. \tag{1}
$$

This system has two undamped natural frequencies of oscillation that satisfy the equation[19]

$$
\omega_n^2 = \frac{m_1(k_1+k_2) + m_2(k_0+k_1)}{2m_1m_2}
$$
$$
\pm \sqrt{\left(\frac{m_1(k_1+k_2) + m_2(k_0+k_1)}{4m_1m_2}\right)^2 - \frac{(k_1+k_2)(k_0+k_1) - k_1^2}{m_1m_2}}. \tag{2}
$$

If components are selected such that $k_0 \gg k_1 \gg k_2$ and $m_2 \gg m_1$, the two resonances are effectively decoupled. The first of these occurs at $\omega_{n1} \approx \sqrt{k_1/m_2}$. This is the resonance of the tail mass on the front spring of the sensor. The second resonance, $\omega_{n2} \approx \sqrt{k_0/m_1}$, is the resonance of the accelerometer mass intimately coupled to the ground's stiffness. The tail mass is effectively stationary for $\omega \gg \omega_{n1}$ where $x_2/x_1 \approx (\omega_{n1}/\omega)(k_1m_1/k_0m_2)$ and the dynamic load on the ground is the mass m_1. Under these circumstances $x_1 \approx x_0$ (the requirement for sensor fidelity) provided that $\omega \ll \omega_{n2}$. Thus the operational bandwidth of an undamped sensor would be $\sqrt{(k_1/m_2)} \ll \omega \ll \sqrt{(k_0/m_1)}$.

The values of the parameters of the model were determined by direct measurement with the exception of the sensor damping (η_1), which was assumed to be small but non-negligible so that the lower resonance would have a reasonable quality factor. Figure 1 shows the predicted response of a ground-contacting sensor design based on measured parameters, which are indicated in the figure. The lower frequency resonance is seen to have very little effect on the output variable (x_1). This small effect would be diminished further if η_1 were assumed to be larger, but it is not significant even for $\eta_1=0$. For the selected value $[\eta_1 = 1.6$ N/(m/s)$]$ the quality factor of the lower resonance is 10, which is higher than experimental measurements. The lower resonance frequency for the selected parameters is 18.5 Hz, which is well below the band of interest. The higher resonance, which has a greater impact on the sensor response, occurs at 3 kHz for the selected parameters. The second resonance is heavily damped (the head mass is about twice the mass corresponding to a critically damped system), and therefore has greater effect on the sensor's phase response than on its amplitude response at frequencies of interest. Below 1 kHz, this amounts to less than a 3° phase delay in the sensor response.

Two ground-contacting sensor prototypes that can be modeled by Eq. (1) were designed, built, and tested. These are shown in Fig. 2. The first design [Fig. 2(A)] a spring-biased sensor, was coupled to the backing structure using a long coil spring with a stiffness of 1.3×10^2 N/m. This permitted the normal force on the ground's surface to be easily varied because this biasing force was the sum of the weight of a 40 g tail mass (0.4 N) and the force exerted by the spring's compression. The sensor head weighed 5 g. It was fabricated with a 0.75 g miniature accelerometer (Vibrametrics model VMI9002) encapsulated in rigid foam. The accelerometer sensitivity and noise density were specified by the manufacturer as 100 mV/g and 10 μg/$\sqrt{\text{Hz}}$.

The second sensor [Fig. 2(B)] was a modular gravity-biased sensor. It was completely decoupled from its backing structure $(k_2=0)$. This sensor was placed on the ground so that it was free standing and then lifted and repositioned

FIG. 2. Two of the vibrometer designs that were tested: (A) a spring-biased sensor and (B) a gravity-biased sensor.

FIG. 3. Modular ground contact sensor transfer function with respect to unloaded ground acceleration (A) amplitude and (B) phase.

2.1×10^3 N/m for the second. In the second design, tubular springs and masses allowed the accelerometers and internal wiring of the sensor to be completely enclosed and weather-proofed.

IV. SENSOR FIDELITY

The fidelity of the ground-contacting sensors described in the previous section was investigated in a series of laboratory experiments using an experimental model filled with damp compacted sand.[20] In the first experiment, surface motion was measured using both a ground contacting sensor (a gravity-biased sensor with a total 156 g tail mass) and a commercial LDV in the same location without the sensor present. The sensor and LDV outputs, corrected for the frequency responses of receiving electronics, were used to generate the transfer function that is plotted in Fig. 3, along with the model prediction for this transfer function. The deviations from the predicted behavior are attributable to several sources, including the calibration of the LDV and the accelerometer that were used, the different areas of the surface that are sampled by each sensor, and the nonlinear elasticity of the sand. Below 1 kHz, the sensor deviates from ideal behavior by less that 20% in amplitude and 10° of phase. Most of the structure above 1 kHz is due to the difference between the footprints of the two sensor types, which made their sensitivities vary for each of the different modes of propagation that were present. This is a complex wavefield that includes several modes of surface-guided and bulk waves in addition to the dominant Rayleigh wave. The model assumes that ground is locally reactive and that both sensors would be equally responsive to all components of the incident signal.

Although the sensor has been modeled as a linear 2-DOF system, sand and soils have shown significant non-

incrementally using a cradle that was not in contact with it during a measurement. The tail mass was a composite made of a 46 cm long 19 mm outer diameter aluminum tube with steel collars attached to it. The long tube permitted the sensor to conform to topography within the array's contact area and the collars allowed for incremental adjustments of the tail mass/biasing force. Each collar was 52 g and the minimum tail mass was 156 g (the mass of the tube with a single collar). The first sensor was one of a kind. The second sensor was built in a quantity of 50 and was designed for use in much larger arrays. The accelerometer for this sensor was selected based primarily on cost and size rather than sensitivity or mass. Here an Analog Devices ADXL103 iMEMs accelerometer was chosen at a cost of less than 1/20th that of the other accelerometers that had been used. This accelerometer had a specified sensitivity and noise density of 1 V/g and 110 μg/$\sqrt{\text{Hz}}$, respectively.

In both sensor designs, the spring represented by k_1 was a section of foam pipe insulation. This permitted the sensor head to move laterally with respect to the tail mass, and to rotate about any axis so that the sensor could be coupled to rough or inclined surfaces and it would not constrain horizontal motion of the ground. This was necessary because the surface motion associated with a Rayleigh wave is elliptical in nature,[12] even though the sensor was only intended to measure its vertical component. The compressive stiffness of the foam springs used in these sensors was measured on a load frame to be 5.4×10^3 N/m for the first design, and

FIG. 4. Accelerations measured from linear scans of a spring-biased sensor away from a seismic source with biasing forces of (A) 1.1 N, (B) 3.6 N, and (C) 11.4 N.

ferences caused by ambient noise. The differences between these plots, which share a common vertical scale, are therefore an indication of nonlinearity. In Fig. 4(A), a 1.1 N biasing force is seen to be sufficient to detect the incident wave and the presence of the buried mine. There are small differences (other than delays) in these waveforms that are indicative of small changes in the sensor's coupling from point to point. In Figs. 4(B) and 4(C), the biasing force is increased to 3.6 N and 11 N, respectively. Here the reproducibility can be seen to improve with the increased biasing force, but the amplitudes of the measured responses change because of the nonlinearity of the sand. For this effect to pose a problem with respect to landmine detection, it would have to degrade the contrast between the mine and the background in addition to reducing the measured response. This is, of course, a function of the algorithm that is used to convert the measured data into an image. Using imaging techniques that have been reported in previous papers, the data shown in Figs. 4(B) and 4(C) will yield nominally identical contrast after processing.[20] With these techniques, the data collected with the 1.1 N biasing force yield a contrast that is degraded by 6–12 dB with regard to the other data sets as a consequence of the reproducibility of the sensor's coupling, rather than the mean fidelity of the measurements. Since this was a synthetic array measurement, the possibility that nonlinearity might adversely affect sensor-to-sensor coupling could not be considered in this analysis.

In another fidelity-related experiment a 20 g sensor head was constructed using a triaxial accelerometer (PCB model W356CA12), with sensitivity and intrinsic noise density comparable to the VMI 9002 accelerometer. This was attached to the spring-biased suspension and used to measure the three components of acceleration on the surface of the damp compacted sand in the experimental model. Figure 5 shows the displacements in the surface normal (Z) and propagation (X) directions that were generated by integration of measurements made 1.5 m from the seismic source. The experiment is configured so that lateral (Y) displacements are negligible in comparison to the other two displacement components. In Fig. 5(A), measurements with a 1.1 N biasing force show that the motion at this location is dominated by two temporally separate pulses. The first of these is characterized by prograde elliptical motion and is consistent with a leaky surface wave (head wave).[24–26] The second arrival is a larger pulse characterized by retrograde elliptical motion that is consistent with a Rayleigh wave. Although a ground truth for this measurement is not available, the measurement is a reasonable match to measurements made with shallow-buried accelerometers in the same model.[23] The measurement indicates that the sensor has not inordinately constrained in-plane motions at the measurement point. In Fig. 5(B) the same measurement is depicted with a 6.2 N biasing force. It is interesting to note here that the nonlinearities apparent in measurements of the surface normal motion do not equally affect both dominant wave types. The motion associated with the leaky wave, which is mediated by compression, is barely altered by the increased biasing force. The

linear behavior.[21–23] In terms of sensor function, this can manifest itself in a modification of the complex stiffness of the ground at the measurement point (k_0 and η_0) as a function of the applied force at that point. To test this, a series of measurements was made using the spring-biased sensor. The biasing force was progressively increased through the compression of the linear coil spring (k_2), while all the dynamic characteristics of the sensor remained unchanged. The data from these measurements are depicted in the waterfall plots shown in Fig. 4. Each plot here represents 101 measurements made 1 cm apart as the sensor is moved away from the seismic source. The positively sloped diagonal connection of the wavefronts of the largest arrivals in these plots represents the incident Rayleigh wave. The enhanced response 50 cm from the source, and backward propagating waves that originate there, are due to a TS50 AP mine that was buried 1 cm below the surface. If the sensor and medium were perfectly linear, each data set would have been identical aside from the dif-

FIG. 5. Hodograms of the displacement measured by a triaxial spring-biased sensor with (A) with a 1.1 N biasing force and (B) with a 6.2 N biasing force.

motion associated with the Rayleigh wave, which is mediated by shear, shows a significant change in both polarity and amplitude at the measurement point.

Although it has not yet been tested experimentally, in-plane displacement information that can be sampled with the ground-contacting sensor, but was not previously available with the radar-based vibrometer, may be exploitable for enhancing the contrast of landmine images. Smith *et al.* have

demonstrated the use of rotational polarity filters to enhance beamforming in a seismic landmine detection system based on pulse echo techniques.[3] Although these techniques have not been extended a full wave-field-imaging seismic system, rotational polarity may offer an additional cue to landmine trigger motion.

V. GROUND COUPLING

As mentioned in the previous section, the reproducibility of ground coupling between sensors and measurement locations appears to be a more significant operational problem than the mean fidelity of measurements. Figure 4 clearly demonstrates that reproducibility is directly related to biasing force; however Figs. 4 and 5 demonstrate that there is a cost in terms of fidelity to arbitrarily increasing this. Although the lost fidelity may not pose a problem for landmine imaging (neglecting a possible effect on mutual scattering), increased biasing force is a potential problem in terms of risk to a system operator. For the data shown in Fig. 3, for example, the largest biasing force that was used (11.4 N) could be sufficient to detonate some types of AP landmines, whereas the smallest allowed a safety factor of at least 10.[14] There is also a cost associated with increasing biasing force in a large array, because the backing structure must support the total load associated with the biasing force. A 1000 sensor array requires a structure with lifting capacity of at least 0.1 ton per N of sensor biasing force. For these practical reasons, it is desirable to achieve reproducible coupling with a minimal biasing force.

In the laboratory model, a variety of contact feet and biasing forces were tested in different sensor designs. A subset of three of these contact feet were tested with several biasing forces in a field experiment to see if one of these permitted acceptable point-to-point reproducibility with a lower biasing force than the others. The three designs that were tested were a flat 1.6 cm diameter disk of Sorbathane™ elastomer, a 90° right-circular polycarbonate cone, and a 60° right-circular polycarbonate cone. The three foot designs were tested at a field site located at a government test facility over a simulated dirt road surface, a gravel-covered roadbed, and a grass-covered surface of comparatively loose dry soil. Data acquired using a 60° conical foot with a 5.6 N biasing force on each these surfaces are depicted in Fig. 6. Each of the data sets shown in this figure represents a linear scan over an area with no known buried targets. It is clear from the figure that a surface wave is sensed for each of the conditions. A common vertical scale normalized by the source's driving current is used for all three plots (the absolute acceleration is irrelevant because a pulse-compression scheme was employed for the measurement[27]). Thus, for a given driving force, which is directly related to drive current for the electrodynamic source, the greatest acceleration response is seen in the gravel roadbed [Fig. 6(B)] and the least response in the grassy area [Fig. 6(C)]. Although the input impedance of the roadbed is higher than the impedance in the grassy area, poor soil consolidation results in more propagation loss and source-coupling nonlinearity in the grass. Similarly, the dirt roadbed [Fig. 6(A)] is a midpoint in
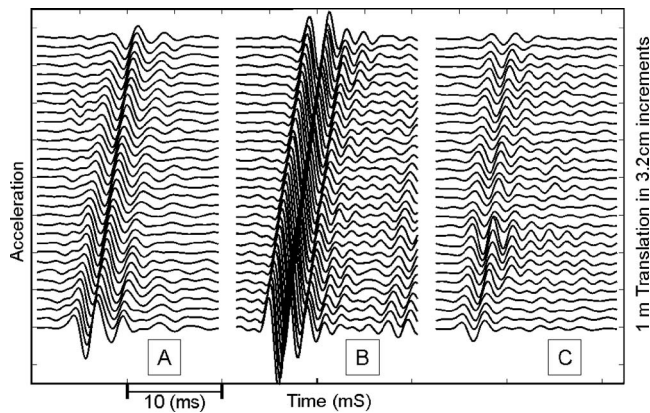
FIG. 6. Windowed waterfall graphs of measured accelerations in a 1 m linear scan away from a seismic at a field site in (A) a dirt roadbed, (B) a gravel roadbed, and (C) a grass-covered area.

FIG. 7. Transfer function of a single ground-contacting sensor in rectangular array/alone (A) amplitude and (B) Phase. Source and sensor locations are shown in inset.

terms of consolidation since it contains finer particles than the grass's sandy soil, and larger rounder particles than the crushed granite base of the gravel roadbed. It is clear from the figure that the soft soil and grass pose a greater reproducibility problem than the rough surface and loose stones of the gravel roadbed. All three of the foot designs were tested with biasing forces down to 1.5 N (the minimum permitted by the hardware configuration). This biasing force proved sufficient in the dirt and gravel roadbeds. Unfortunately, the test range that was used did not include either grass-covered landmine burials or grass-covered regions of sifted soil. Given the history of the site, it is likely that both manmade and natural scatterers (tree roots, rocks, etc.) were present in the grassy area. Thus a direct comparison to the other two carefully controlled soil conditions is difficult.

VI. MUTUAL SCATTERING

Although the ground-contacting sensors were designed to be suitable for large physical arrays, these arrays pose a potential problem as compared to noncontact sensor arrays in that the motions of individual sensors may couple through the ground and impede the function of the array. To thoroughly test coupling effects experimentally would require a very large number of sensors or suitable dynamic surrogates. Because these were not available, limited testing of interaction effects was performed using the sensors that were constructed for the field test.

The mutual scattering experiment involved a rectangular array of 10 sensors in the direction of the source (X), and 3 sensors parallel to the incident wavefronts (Y). The spacing of these elements was 3.4 cm × 10 cm in each direction, respectively. In the experiment, the array was first populated by a single centrally located sensor positioned 1.5 m from the seismic source in the laboratory experimental model. The remaining elements of the array were then populated without disturbing the first sensor. The array's transfer function was constructed as the ratio of the first sensor's response to a seismic incident signal within the array to that sensor's response when the other array locations were empty. This transfer function is plotted in Fig. 7. From the figure, it is apparent that between 150 Hz and 1100 Hz the surrounding
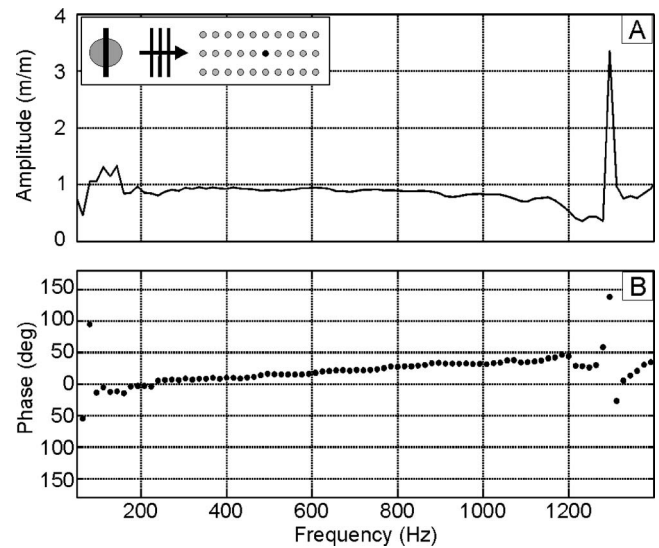
array reduces the output of the sensor slightly ($\lesssim 3$ dB). A small amplification below 150 Hz appears to be indicative of changes in the background noise, which was produced by machinery in adjoining laboratories. The spike at ~1300 Hz is due to a slight downward shift in a spectral null that was caused by interference of the various modes and paths comprising the received signal when the array was populated. The phase of the transfer function shows that there is a delay (linear frequency-phase relationship) between the two measurements. This delay is approximately 14 μs, with the sensor-in-array output leading that of the lone sensor. A plausible explanation for this is that there is a local increase in the shear modulus of the sand around each sensor that is caused by the applied biasing force. This is consistent with the observed reduction in sensor response that results from increasing the biasing force. If this delay is equated with a change in propagation speed over the portion of the array that lies between the sensor and the source (5 sensors spanning 17 cm), it predicts that the array increases the incident signal's mean propagation speed (dominantly the Rayleigh wave speed, although other modes are present) by approximately 1% over its extent. If the problem is viewed as a one-dimensional transmission line, the propagation speed change can only account for a 0.5% amplitude reduction in signal level. More likely explanations for the reduced sensor output observed following the introduction of the array are that it is due to loss in the foam springs of the intervening sensors, or due to shadowing.

VII. LANDMINE IMAGING

The field experiment that was done to evaluate sensor performance involved a linear array of 32 gravity-biased sensors with 156 g tail masses spaced 3.2 cm apart. The array was scanned in 32 steps perpendicular to its alignment to form a semisynthetic 1 m^2 planar array. Figure 8 shows the experimental setup that was used at the field site. The seismic source was oriented both for a broadside and for an
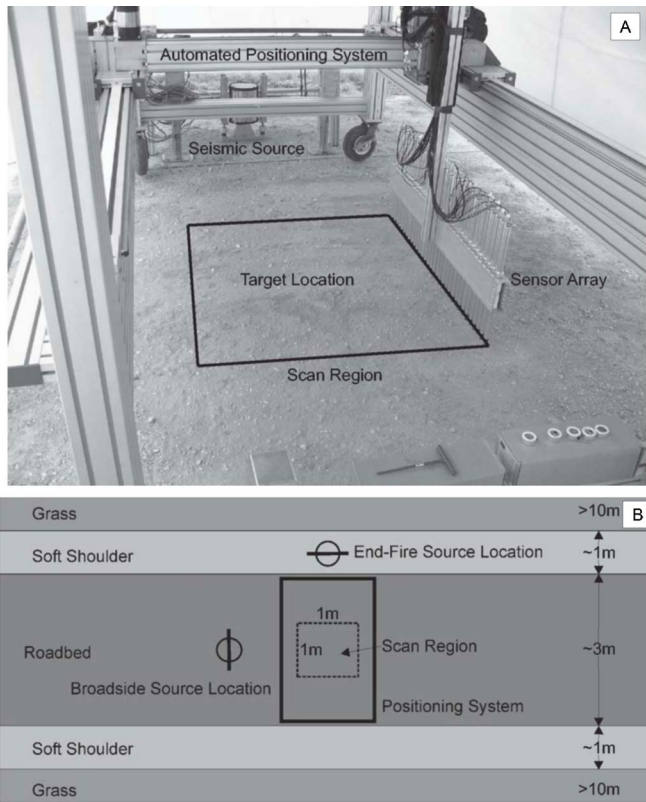
FIG. 8. Configuration of field experiment: (A) photograph of system on simulated dirt roadbed and (B) layout of experiment at field site.

end-fire illumination of the linear array. Although this should have provided insight into mutual scattering effects, the roadbeds at the test site exhibited significant horizontal anisotropy, which was the dominant difference between the two incident signal aspects. There was a horizontal gradient in surface wave speed of the roadbed, which slowed near both shoulders. For all of the measurements, the linear array was located perpendicular to the roadbed. Thus, the end-fire measurements required the source to be located in the low-speed region.

Processed images from the field measurements for various landmine types, road conditions, and propagation directions are depicted in Figs. 9 and 10. Figure 9 contains images formed with data collected in a dirt roadbed, and Fig. 10 shows images from an adjoining gravel roadbed. The targets were all AT landmines or landmine simulants that had been well-weathered in place prior to the experiment to ensure that none of the detections were tied to trenching effects. The array was located to place targets in the approximate center of the scan region, but some of the targets' locations were not well-known and these are offset slightly. Collectively, the images demonstrate sufficient reproducibility and fidelity across the array for the conditions that were examined. This is remarkable given the size of the largest stones on the gravel roadbed, which was comparable to the diameter of the sensor. At many measurement locations, sensors were in contact with a single stone, and sensors frequently sat at steep angles of repose atop these stones. Due to the anisotropy of the roadbeds, there are obvious differences in the images
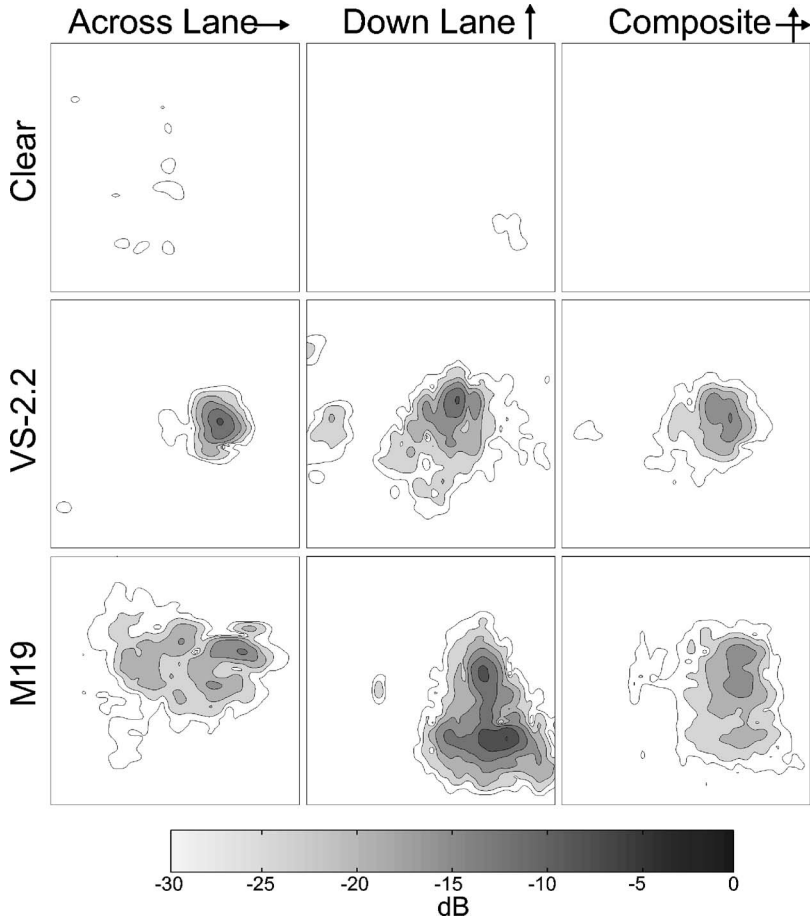
FIG. 9. Images of AT landmines in a dirt roadbed at the field test site: Rows represent mine types as labeled and columns represent the propagation direction of the interrogation signal with he roadbed as labeled. Color is scaled to the maximum pixel value (M19, down lane).

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Martin *et al.*: Seismic landmine detection   2683
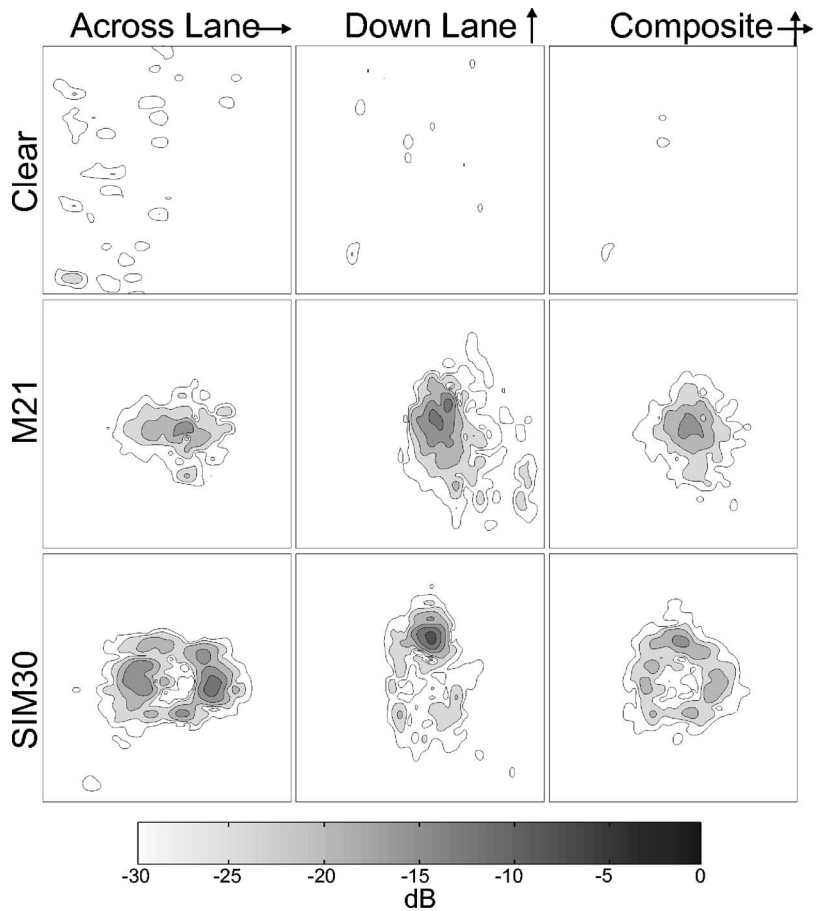
FIG. 10. Images of AT landmines in a gravel roadbed at the field-test site: Rows represent mine types as labeled and columns represent the propagation direction of the interrogation signal with respect to the roadbed as labeled. Color is scaled to the maximum pixel value from Fig. 9.

formed using different source locations in otherwise identical scans. Images formed from incident seismic signals that propagated perpendicular to the road bed (shown in the first columns of Figs. 9 and 10) generally show less background contrast than those that were formed from incident signals propagating parallel to it (shown in the middle columns of the figures).

All of the data were processed with an imaging algorithm that had been developed and refined for use with synthetic array measurements employing a radar-based sensor. The algorithm does not account for the spreading and attenuation of the incident signal and is therefore more sensitive to objects closer to the source. The algorithm may also smear large targets in the direction of the source. Thus, clutter in the original data, regardless of its cause, will produce image clutter that is dependent on its location relative to both the seismic source and the buried target.

In a system designed to search progressively through a region, the deficiencies of the imaging algorithm would need to be accepted or remedied through the use of alternative signal processing techniques. The confirmation-system application that was previously discussed permits this problem to be addressed through a reconfiguration of system hardware. Since the region to be inspected has been prescribed on all sides for the confirmation system, it may be illuminated from any direction or from multiple directions. Each different look direction will result in a shift of the artifact-dominated contributions to the background, while the location of buried objects will be fixed in space. This is apparent in the corresponding pairs of images in the first and second columns of Figs. 9 and 10. It should therefore be possible to improve resolution in the resulting image by combining data from several look directions. This can be done without adding to measurement time by exploiting spread spectrum techniques and simultaneously exciting all of the sources with coded signals.[28,29]

The potential to improve image resolution and contrast with multiple look directions was tested with data collected in the field experiment because the laboratory model is not large enough to permit this sort of experiment. The images shown in the first two columns of Figs. 9 and 10 were combined to form the corresponding images in the third column. The images were combined by taking the square root of the product of corresponding image pixel values. The combined images show better spatial resolution than either of their precedents. The reduction in the overall background ($\sim$3 dB) is not as striking as the reduction of the peak levels measured over some of the landmines. Improved spatial resolution is particularly apparent for the data collected over the SIM30 landmine simulant in the gravel roadbed. Here the combined image clearly outlines the perimeter of the simulant. Any number of data sets collected using different source locations could be integrated in this way, although there are other schemes that can be used to combine these data sets, which may offer advantages over the chosen technique.

VIII. CONCLUSIONS

Ground contacting sensors have been designed, built and tested for use in a seismic landmine-detection system. The sensor design that has been considered is not as easily scanned as noncontacting sensors are, but is suited to the incremental motion of a detection-confirmation system. Its deployment requires that a psychological barrier to touching the ground above a buried mine be overcome. Experimental testing indicates that the current design is capable of meeting reasonable requirements for cost, sensitivity, fidelity, nonintrusion, reproducibility of ground coupling, and large-array compatibility. Measurements that have been conducted with these sensors in both a laboratory model and at a realistic outdoor site reveal that the most significant deviations from ideal behavior are attributable to the nonlinearity of the soil that a sensor contacts.

[1] J. C. Cook and J. J. Wormser, "Semi-remote acoustic, electric and thermal sensing of small buried nonmetallic objects," IEEE Trans. Geosci. Electron. **GE-11**(3), 135–152 (1973).

[2] "Feasibility of acoustic landmine detection: Final technical report," BBN Technical Report No. 7677, May 1992.

[3] E. Smith, P. S. Wilson, F. W. Bacon, J. F. Manning, J. A. Behrens, and T. G. Muir, "Measurement and localization of interface wave reflections from a buried target," J. Acoust. Soc. Am. **103**, 2333–2343 (1998).

[4] J. M. Sabatier and N. Xiang, "An investigation of acoustic-to-seismic coupling to detect buried antitank landmines," IEEE Trans. Geosci. Remote Sens. **39**, 1146–1154 (2001).

[5] W. R. Scott, Jr., J. S. Martin, and G. D. Larson, "Experimental model for a seismic landmine detection system," IEEE Trans. Geosci. Remote Sens. **39**, 1155–1164 (2001).

[6] K. Sarabandi and D. E. Lawrence, "Acoustic and electromagnetic wave interaction: Estimation of Doppler spectrum from an acoustically vibrated metallic circular cylinder," IEEE Trans. Antennas Propag. **51**, 1499–1507 (2003).

[7] N. Xiang and J. M. Sabatier, "An experimental study on antipersonnel landmine detection using acoustic-to-seismic coupling," J. Acoust. Soc. Am. **113**, 1333–1341 (2003).

[8] D. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, "Nonlinear seismo-acoustic land mine detection and discrimination," J. Acoust. Soc. Am. **111**, 2705–2714 (2002).

[9] S. H. Lee and W. R. Scott, Jr., "A focused radar antenna for use in seismic mine detection systems," Radio Sci. **39** (2004).

[10] J. S. Martin, D. J. Fenneman, F. Codron, P. H. Rogers, W. R. Scott, G. D. Larson, and G. S. McCall, "Ultrasonic displacement sensor for the seismic detection of buried land mines," in *Proceeding SPIE: 2002 Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, Orland, FL, April 2002, Vol. **4742**, pp. 606–616.

[11] A. G. Petculescu and J. M. Sabatier, "Air-coupled ultrasonic sensing of grass-covered vibrating surfaces; qualitative comparisons with laser Doppler vibrometry," J. Acoust. Soc. Am. **115**, 1557–1564 (2004).

[12] J. D. Achenbach, *Wave Propagation in Elastic Solids*, 7th ed. (Elsevier Science, New York, 1993), pp. 187–194.

[13] A. Behboodian, W. R. Scott, Jr., and J. H. McClellan, "Signal processing of elastic surface waves for localizing buried land mines," in *Proceedings of the 33rd Assilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, **2 pt.2**, pp. 827–830 (1999).

[14] *Jane's Mines and Mine Clearance*, 9th ed. edited by C. King (Janes Information Group, Alexandria, VA, 2004), pp. 63–463, and 195, 199, 265, 269, and 357 contain salient examples of pressure-fused AP mine triggering forces.

[15] R. Bachrach and A. Nur, "High-resolution shallow-seismic experiments in sand, Part I: Water table, fluid flow, and saturation," Geophysics **63**(4), 1225–1233 (1988).

[16] R. Bachrach, J. Dvorkin, and A. Nur, "High-resolution shallow-seismic experiments in sand, Part II: Velocities in shallow unconsolidated sand," Geophysics **63**(4), 1234–1240 (1998).

[17] M. Westebbe, J. F. Bohme, H. Krummel, and M. B. Matthews, "Model fitting and testing in near surface seismics using maximum likelihood in frequency domain," in *Conference Record 32rd Asilomar Conference Signals, Systems, and Computers*, Pacific Grove, CA, Vol. 2, pp. 1311–1315 (1998).

[18] J. M. Sabatier, H. E. Bass, L. N. Bolen, and K. Attenborough, "Acoustically induced seismic waves," J. Acoust. Soc. Am. **80**(2), 646–649 (1986).

[19] L. Meirovitch, *Elements of Vibration Analysis*, 2nd ed. (McGraw-Hill, New York, 1986), pp. 107–116.

[20] J. S. Martin, G. D. Larson, and W. R. Scott, Jr., "Surface-contacting vibrometers for seismic landmine detection," in *Proceedings of SPIE: 2005 Annual International Symposium on Aerospace/Defense Sensing Simulation, and Controls*, Orlando, FL, April, 2005, Vol. **5794**, pp. 590–600.

[21] J. C. Santamarina, K. A. Klein, and M. A. Fam, *Soils and Waves: Particulate Materials Behavior, Characterization, and Process Monitoring* (Wiley, Chichester, England, 2001), pp. 103–115 and 246–247.

[22] M. S. Korman and J. M. Sabatier, "Nonlinear acoustic techniques for landmine detection," J. Acoust. Soc. Am. **116**, pp. 3354–3369 (2004).

[23] G. D. Larson, J. S. Martin, P. H. Rogers, W. R. Scott, and G. S. McCall, "Impact of nonlinear effects on high-frequency seismic landmine detection," in *Proceedings of the 5th International Symposium on Technology and the Mine Problem*, April 22–25 2002, Monterey California (http://www.hdic.imu.edu/dtif/Conferences/Monterev2/LAND/MARTIN. PDF).

[24] L. Tsang, "Time-harmonic solution of the elastic head wave problem incorporating the influence of Rayleigh poles," J. Acoust. Soc. Am. **63**(5), 1302–1309 (1978).

[25] C. T. Schröder and W. R. Scott, Jr., "On the complex conjugate roots of the Rayleigh equation: The leaky surface wave," J. Acoust. Soc. Am. **110**(6), 2867–2877 (2001).

[26] J. G. Harris and J. D. Achenback, Comment on "On the complex conjugate roots of the Rayleigh equation: The leaky surface wave," J. Acoust. Soc. Am. **112**(5), 1747–1748 (2002).

[27] J. S. Martin, W. R. Scott, G. D. Larson, P. H. Rogers, and G. S. McCall II, "Probing signal design for seismic landmine detection," in *Proceedings of the SPIE: 2004 Annual International Symposium on Aerospace/Defense Sensing, Simulation, and Controls*, Orlando, FL, April 2004, Vol. **5415**, pp. 133–144.

[28] H. L. Van Trees, *Detection, Estimation, and Modulation Theory Part I* (Wiley, New York, 1968), pp. 257–287 and 381–386.

[29] A. J. Viterbi, *CDMA, Principles of Spread Spectrum Communication* (Addison-Wesley, Reading, MA, 1995), pp. 4–51.

Analysis of normal incidence absorption of pyramidal polyurethane foam by three-dimensional finite element frequency domain acoustical analysis

Huoy-Shyi Tsay[a)]

Department of Mechanical and Electro-Mechanical Engineering, Tamkang University,
151 Ying-Chuan Road, Tamsui, Taipei Hsien, Taiwan 25137, Republic of China

A general way to enhance the sound absorption capability of polyurethane foam is the use of a nonplanar surface such as the pyramidal surface. The use of the pyramidal surface can greatly enhance the sound absorption of the foam in the mid and high frequency ranges. In this study, a three-dimensional finite element frequency domain acoustical analysis is used to analyze the normal incidence sound absorption of the pyramidal polyurethane foam. Such a numerical method is more efficient than the earlier studies by saving the calculation time on the elements of the incident acoustic field. In the study, the maximum broadband averaged normal incidence sound absorption of the pyramidal polyurethane foam is pursued based on the change of the pyramidal shape. The variation of the pyramidal shape is performed by changing the angle between two sides of a pyramid but keeping the pyramid base and the foam volume the same throughout the study. It is found that the pyramidal polyurethane foam with pyramidal angle of 29° has the highest broadband averaged normal incidence sound absorption in the 100–4k, 500–4k, as well as 2k–4k Hz ranges. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354044]

I. INTRODUCTION

Upon the enhancement of the absorption coefficient in the desired frequency band, foams with complex other than planar surface are usually produced. Conventionally, acoustical properties of foams are numerically analyzed by applying the time-domain porous finite elements coupled with finite elements of an incident acoustic field with which the foam is in contact. The transmission and reflection waves that characterized acoustical properties are then solved. Among these analyses, two major constituent models for describing acoustical properties of foams can be found. First, the foam is seen as an isotropic mixed fluid and the propagation of sound is determined by the empirical power law expressions of air resistance divided by frequency[1] or other physical parameters.[2] In the other aspect, the acoustical properties of the foam are described by the Biot's theory of poroelasticity.[3–5]

Based on the isotropic mixed fluid constituent model, Craggs[6,7] analyzed a one-dimensional, plane or axisymmetric acoustic absorption system for obtaining the absorption coefficient. Easwaran and Munjal[8] predicted the reflection characteristics of rectangular and conical wedges in an impedance tube. Later, Chen *et al.*[9] investigated the absorption coefficient of corrugated porous materials with triangle, semicircle, and convex rectangular surface shapes. There is only one longitudinal wave to be analyzed, if the foam is presumed as an isotropic mixed fluid. To design the acoustical properties of foams, methods that combine the acoustic field with the foam modeling by the Biot's poroelastic theory are more effective. Since, behaviors of two longitudinal waves and one shear wave of the foam can be analyzed. Using the Biot's theory, Kang and Bolton[10] studied the optimal design of axisymmetric foam wedges. Heretofore, the additional acoustic field is needed to perform simulations.

In order to eliminate the use of the incident acoustic field, Tsay and Yeh[11] applied the Biot's theory and used a normal incident impulse pressure excitation to actuate directly on the upper surface of a rigidly backed planar foam. After applying this new approach, the exact acoustic property solutions of planar foams with permeable or impermeable upper surfaces were successfully derived. Later, this approach was extended and the finite element frequency domain acoustical analyses with triangular, four-node quadrilateral, and four-node tetrahedral porous elements were formulated.[12,13] Based on the newly developed finite element frequency domain methods, the acoustical properties of planar polyurethane foams were analyzed. Subsequently, the influence of circumferential edge constraint on the acoustical properties of polyurethane foam samples was examined.

In this study, the three-dimensional finite element frequency domain acoustical analysis with tetrahedral porous elements is applied to analyze the normal incidence sound absorption of the pyramidal polyurethane foams. In the following sections, the three-dimensional finite element frequency domain acoustical analysis and material properties are briefly introduced, and the pursuing of pyramidal shape of the maximum broadband averaged normal incidence sound absorption based on the constraints of constant foam base and constant foam volume is studied.

---

a)Electronic mail: huoyshyi@mail.tku.edu.tw
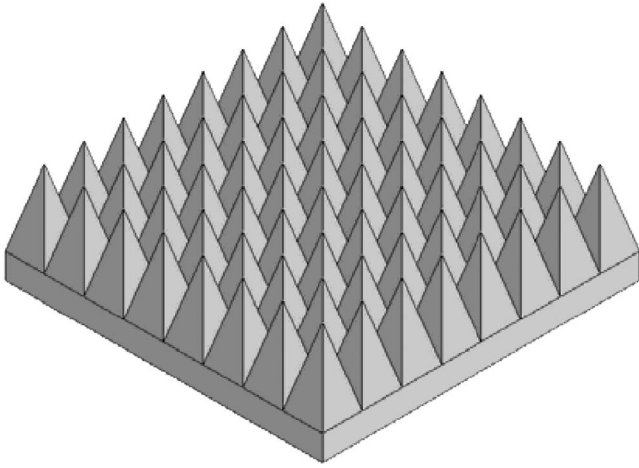
FIG. 1. A piece of pyramidal polyurethane foam.

II. THREE-DIMENSIONAL FINITE ELEMENT FREQUENCY DOMAIN ACOUSTICAL ANALYSIS

A. Finite element model

The acoustical properties of the pyramidal polyurethane foam shown in Fig. 1 with it base fixed on an impermeable rigid plane are investigated in this study. One section of the pyramidal polyurethane foam is presented in Fig. 2. In Fig. 2, L and W are the length and width of the pyramid base, H is the height measured from the base to the tip, H_s is the height of the side edge, θ is the angle between two sides of a pyramid, and $p_0\delta(t)$ is the normal incident impulsive sound pressure loading of amplitude p_0. It is noted that the pyramidal polyurethane foams considered in this study are assumed to be stressed from the initially un-deformed position. Hence, the initial conditions for the foam are $u_i=0$, $U_i=0$, $\dot{u}_i=0$, and $\dot{U}_i=0$ ($i=x,y,z$) at $t=0$. u_i and U_i ($i=x,y,z$) are displacement components of the solid skeleton and the fluid, respectively. The foam is stress free at the initial state, thus, all stresses are zero at $t=0$. According to the Biot's poroelastic theory and the Cartesian coordinates shown, the dynamic equations for the three-dimensional foam problems after Laplace transform with respect to the time variable t can be written in a compact form[13] as

FIG. 3. Four-node tetrahedral porous element.

$$N\nabla^2\bar{u}_i + \nabla[(A+N)\bar{u}_{i,i} + Q\bar{U}_{i,i}]$$
$$- s^2(\rho_{11}\bar{u}_i + \rho_{12}\bar{U}_i) - bs(\bar{u}_i - \bar{U}_i) = 0, \tag{1}$$

$$\nabla(Q\bar{u}_{i,i} + R\bar{U}_{i,i}) - s^2(\rho_{12}\bar{u}_i + \rho_{22}\bar{U}_i) + bs(\bar{u}_i - \bar{U}_i) = 0, \tag{2}$$

where N, A, Q, and R are elastic constants; b is the dissipation coefficient; ρ_{11}, ρ_{22} and ρ_{12} are apparent mass densities. ∇^2 is the Laplacian and ∇ is the gradient. $\bar{u}_i$ and $\bar{U}_i$ are the Laplace transforms of u_i and U_i ($i=x,y,z$). s is the Laplace transform parameter.

In the previous study,[13] the Galerkin finite element method with four-node tetrahedral porous elements was first applied to obtain acoustical properties of the planar polyurethane foam in Laplace domain. Figure 3 shows a four-node tetrahedral porous element. Upon considering three Cartesian coordinate components of displacements of the solid skeleton and of the fluid at each node, the column vector of displacements of a four-node tetrahedral porous element with 24 degrees of freedom was presented as $q=[q_1,\ldots,q_{24}]^T$. In Fig. 3, $P(x,y,z)$ represents an arbitrary point within an element. In terms of a natural coordinate system (ξ,η,ζ), the tetrahedral porous element defined with reference to the Cartesian coordinate system (x,y,z) was transformed into a master element as shown in Fig. 4. $P(x,y,z)$ after the coordinate system transformation was expressed as $P(\xi,\eta,\zeta)$. According to the natural coordinates defined, four Lagrange-type shape functions for the tetrahedral porous element were represented as

FIG. 2. Pyramidal polyurethane foam with its upper surface subjected to a normal incident impulsive sound pressure loading.
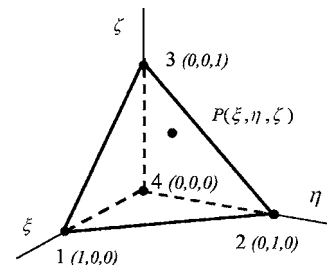
FIG. 4. Four-node master element in ξ, η, and ζ space.

$$M_1 = \xi, \; M_2 = \eta, \; M_3 = \zeta, \; M_4 = 1 - \xi - \eta - \zeta. \qquad (3)$$

With the use of the column vector of displacements $\boldsymbol{q}$ and the shape functions defined, the solid skeleton and fluid displacement components were further written in a matrix form as

$$\hat{\boldsymbol{u}} = \boldsymbol{M}\boldsymbol{q}, \qquad (4)$$

where $\hat{\boldsymbol{u}} = [\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z]^T$ and $\boldsymbol{M}$ is the shape function matrix, which is shown in Appendix A. Following the Galerkin approach, the equation governing the dynamic behavior of the tetrahedral porous element was derived, which is presented in Appendix B. After the substitution of shape functions [Eq. (3)] and displacement components [Eq. (4)], the dynamic equation of the tetrahedral porous element [Eq. (B1)] was simplified as

$$\sum_{e=1}^{n} [\boldsymbol{K}^{(e)}\boldsymbol{q} - \boldsymbol{F}^{(e)}] = 0, \qquad (5)$$

where $\boldsymbol{K}^{(e)}$ is the stiffness coefficient matrix for the element and $\boldsymbol{F}^{(e)}$ is viewed as the loading vector which contains boundary conditions. A section of the pyramidal polyurethane foam can be analyzed by discretizing it with tetrahedral porous elements. The individual element equations would then be assembled to form the equations for the pyramidal polyurethane foam. After manipulations, the Laplace domain finite element equation for the pyramidal polyurethane foam can be obtained as

$$\boldsymbol{K}^* \boldsymbol{q}^* = \boldsymbol{F}^*, \qquad (6)$$

where $\boldsymbol{K}^*$ is the global stiffness matrix, $\boldsymbol{q}^*$ is the global displacement vector, and $\boldsymbol{F}^*$ is the global loading vector. The Laplace transformed dynamic stiffness $\bar{K}(s)$ evaluated from the top surface of the pyramidal polyurethane foam for normal incidence case can then be defined as $\bar{K}(s) = -p_0 / [(1-\phi)\bar{u}_z + \phi\bar{U}_z]$. Once the Laplace transformed dynamic stiffness $\bar{K}$ is obtained, the Laplace transformed surface acoustic impedance $\bar{Z}$, the reflection coefficient R_0, and the sound absorption coefficient A_0 can be formulated as

$$\bar{Z} = \frac{\bar{K}(s)}{s}, \qquad (7)$$

$$R_0 = \frac{\bar{Z} - Z_{\text{air}}}{\bar{Z} + Z_{\text{air}}} = \frac{\bar{K}(s)/s - Z_{\text{air}}}{\bar{K}(s)/s + Z_{\text{air}}}, \qquad (8)$$

$$A_0 = 1 - |R_0|^2, \qquad (9)$$

where Z_{air} is the characteristic impedance of air. Frequency response functions of K, Z, R_0, and A_0 can be obtained by replacing the Laplace transform parameter s that appears in Eqs. (7)–(9) by $j\omega$, where ω is the radial frequency.
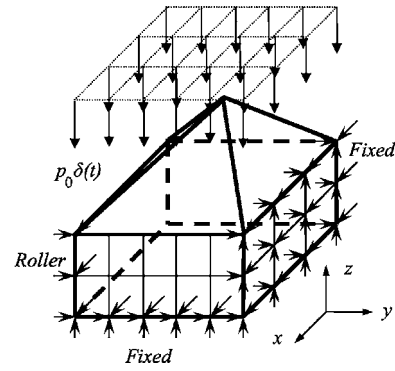
FIG. 5. Boundary conditions of a pyramidal polyurethane foam.

B. Boundary conditions

As shown in Fig. 5, the amplitude of the normal incident impulsive sound pressure loading applied on the top surface of the pyramidal polyurethane foam is presented as $p = p_0\delta(t)$. If the top surface of the pyramidal polyurethane foam is permeable, the solid skeleton and the fluid of the foam, depending on the porosity, share the pressure loading. Thus, the boundary conditions of the solid skeleton and the fluid on the top surface after the Laplace transform can be described as $\bar{\sigma}_{zz} = -(1-\phi)p_0$ and $\bar{\sigma} = -\phi p_0$, respectively. The base of the pyramidal polyurethane foam is presumed to be fixed on an infinite rigid plane. Hence, the Laplace transformed solid skeleton and fluid displacements are zero on the base of the pyramid.

The side edge surfaces of one section of the pyramidal polyurethane foam can be all set as roller condition for simulating a laterally successively extent pyramidal foam, which is shown on the left-hand side edge surface. The roller condition implies that the node can freely move up and down but not to the sides; hence, the Laplace transformed solid skeleton and fluid displacements normal to the side edge surfaces can all be set as zero. For the case that a side edge surfaces of one section of the pyramidal polyurethane foam is glued to a rigid surface, the side edge restraints can be simulated as fixed, as presented on the right-hand side edge surface. Same as the condition of the pyramid base, fixed condition implies that the solid skeleton and fluid displacements are all zero on the edge surface.

III. MATERIAL PROPERTIES OF POLYURETHANE FOAMS

In order to predict the acoustical frequency response functions of the pyramidal polyurethane foam, the apparent mass densities ρ_{11}, ρ_{12}, ρ_{22}, the dissipation coefficient b, and the elastic constants N, A, Q, R can all relate to the foam constituents, porosity ϕ, tortuosity α_∞, pore cross-section shape correction factor c, and direct current (dc) flow resistivity σ_0. The typical material properties of polyurethane foam saturated with air were discussed in detail in the previous study[11] and are summarized as follows.

Based on the tortuosity, the coupling apparent mass ρ_{12} was represented as $-\phi\rho_f(\alpha_\infty-1)$. Subsequently, the total effective mass of the solid and fluid, ρ_{11} and ρ_{22}, were defined as $(1-\phi)\rho_s-\rho_{12}$ and $\phi\rho_f-\rho_{12}$. ρ_f is the mass density of fluid and ρ_s is the mass density of solid. The Laplace transformed dissipation coefficient $\bar{b}$ defined in the dynamic poroelastic theory was reformulated as

$$\bar{b} = \frac{-\sigma_0}{4}\phi^2\left\{\frac{[m_{r_0}J_1(m_{r_0})]/[J_0(m_{r_0})]}{1-[2J_1(m_{r_0})]/[m_{r_0}J_0(m_{r_0})]}\right\}, \tag{10}$$

where J_0 and J_1 are Bessel functions of the first kind of order zero and one. The dc flow resistivity σ_0 was derived as $(8c\mu\alpha_\infty)/(\phi r_0^2)$. μ is the viscosity of air, r_0 is the radius of the pore, and c is a cross-section shape correction factor. Following the definition of the dc flow resistivity σ_0, the collected material coefficient m_{r_0} appearing in Eq. (10) can be rewritten as $m_{r_0}=c\sqrt{-8s\rho_{22}/(\sigma_0\phi^2)}$. Elastic coefficients of the polyurethane foams, A, Q, and R, were formulated as the combinations of porosity ϕ, the shear modulus of the solid skeleton in vacuum N, the bulk modulus of the solid skeleton K_b, and the bulk modulus of the fluid K_f as

$$A = \frac{(1-\phi)^2}{\phi}K_f + K_b - \frac{2}{3}N, \tag{11}$$

$$Q = (1-\phi)K_f, \tag{12}$$

$$R = \phi K_f. \tag{13}$$

For the air saturated polyurethane foams, N, K_b, and K_f are complex coefficients, and K_b was evaluated by

$$K_b = \frac{2N(1+\nu)}{3(1-2\nu)}, \tag{14}$$

where ν is the Poisson's ratio of the foam. The bulk modulus of air K_f was derived using the equation of state of air and the equation of thermal conduction of air. With the cross-section shape correction factor and tortuosity considered, the Laplace transformed bulk modulus of air $\bar{K}_f$ was formulated as

$$\bar{K}_f = \frac{p_{amb}\gamma}{\gamma-(\gamma-1)\{1-2/Bm_{r_0}[J_1(Bm_{r_0})]/[J_0(Bm_{r_0})]\}}, \tag{15}$$

where p_{amb} is the ambient mean pressure, B is the square root of the Prandtl number, and $\gamma = c_p/c_v$. c_v and c_p are specific heats per unit mass at constant volume and pressure, respectively.

IV. INFLUENCE OF PYRAMIDAL SHAPE ON THE NORMAL INCIDENCE SOUND ABSORPTION OF THE PYRAMIDAL POLYURETHANE FOAM

The normal incidence sound absorption results of the planar polyurethane foam of thickness 0.0508 m examined in the previous study[13] are used as comparators to that of the

TABLE I. Material properties of the polyurethane foam (See Ref. 11).

ϕ	Porosity	0.95
N (10^4 N/m^2)	Shear modulus	$55(1+0.2i)$
ν	Poisson coefficient	0.35
σ_o (N s/m^4)	dc flow resistivity	5500
α_∞	Tortuosity	1.0
ρ_s (kg/m^3)	Density of frame	45.7
c	Shape factor	1.0
h (m)	Thickness of foam	0.0508

pyramidal polyurethane foam as the pyramidal shape modified. The material properties of the planar polyurethane foam are presented in Table I and the normal incidence sound absorption coefficient of the foam is shown in Fig. 6. Figure 6 shows that the normal incidence sound absorption coefficients of the foam samples can be accurately predicted by the three-dimensional finite element frequency domain acoustic analysis (3D FEFDAA) with suitable boundary conditions. The broadband averaged normal incidence sound absorption coefficient of the planar polyurethane foam at $1k$–$4k$ Hz range is around 0.87 and there is a possibility for the enhancement of the sound absorption coefficient of the foam through the change of its surface shape from plane to pyramid.

In the following study, the diameter of 0.0395 m of the foam sample is arbitrary chosen as the width and length of the pyramid base and the dimension of the pyramid base is kept the same throughout the study. The influence of the change of pyramidal shape on the acoustical properties of the foam is studied based on the constraint of constant foam volume of $0.0395 \times 0.0395 \times 0.0508$ m^3 and is performed by
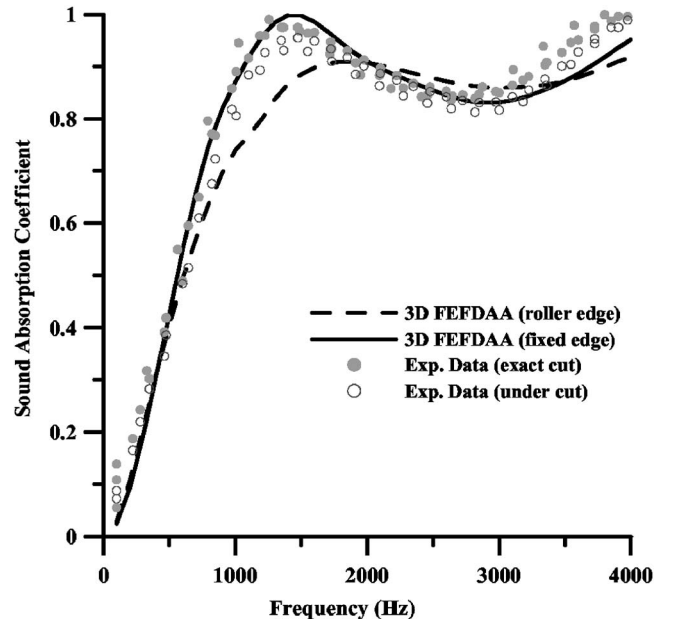
FIG. 6. Normal incidence sound absorption coefficient of the planar polyurethane foam backed with a rigid wall (See Ref. 13) (thickness =0.0508 m, sample diameter=0.0395 m).

TABLE II. Dimensions of pyramidal polyurethane foam models ($W=L$ =0.0395 m, volume=$0.0395 \times 0.0395 \times 0.0508$ m^3=7.9261×10^{-5} m^3).

Model	$\theta°$	H (m)	H_s (m)
1	21	0.1219	$0.3 \times 0.0508=0.0152$
2	24	0.1118	$0.4 \times 0.0508=0.0203$
3	29	0.1016	$0.5 \times 0.0508=0.0254$
4	36	0.0914	$0.6 \times 0.0508=0.0305$
5	47	0.0813	$0.7 \times 0.0508=0.0356$
6	66	0.0711	$0.8 \times 0.0508=0.0406$
7	105	0.0610	$0.9 \times 0.0508=0.0457$
8 (Plate)	180	0.0508	$1.0 \times 0.0508=0.0508$

the change of the angle between two sides of a pyramid that is defined in Fig. 2. The dimensions and the model numbers of the pyramidal polyurethane foams examined are presented in Table II and the nodes, elements, and the tetrahedral porous element models of pyramidal foams used in the 3D FEFDAAs are shown in Fig. 7. In the 3D FEFDAAs, the elements are arranged symmetrically to the x and y axes, and the restraints of the nodes on the side edge surfaces of a pyramid section are assumed as rollers for simulating the laterally successively extent pyramidal foam.

The analyzed results of the normal incidence sound absorption coefficients and the reflection coefficients of the pyramidal polyurethane foams are shown in Figs. 8 and 9, respectively. The analyzed normal incidence sound absorption coefficients and reflection coefficients results clearly show that the change of the upper surface of the foam from plane to pyramid does enhance the normal incidence sound absorption capability of the foam in the mid and high frequency regions ($1k$–$2k$ and $2k$–$4k$ Hz). Furthermore, the reduction of the pyramidal angle gradually increases the normal incidence sound absorption capability of the foam to a certain maximum then decreases its acoustic performance. Figure 10 and Table III show the broadband averaged normal incidence sound absorption coefficients of various pyramidal foams in
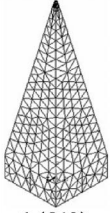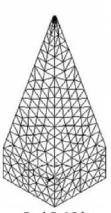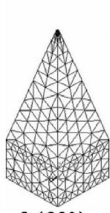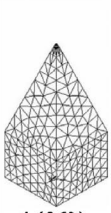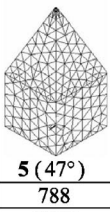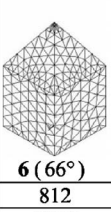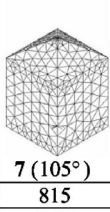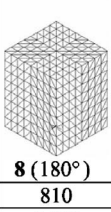
FIG. 7. Nodes, elements, and tetrahedral porous element models of the pyramidal polyurethane foams.
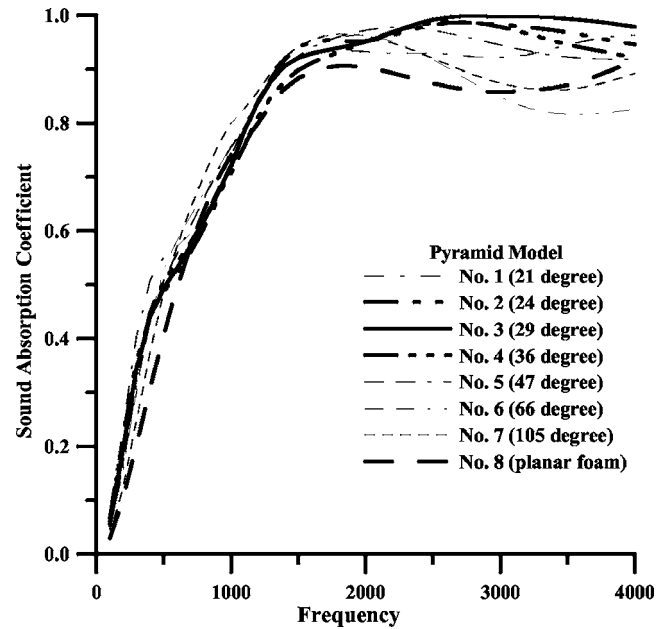
FIG. 8. Influence of pyramidal angle on the normal incidence sound absorption coefficient.
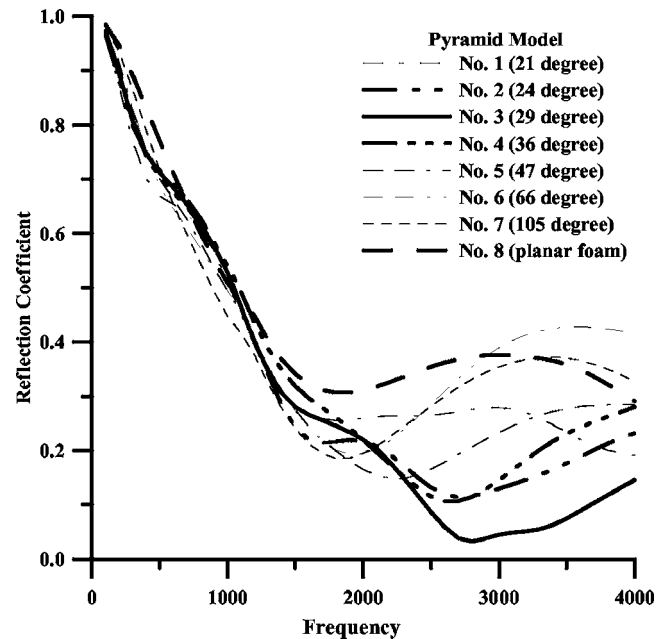
FIG. 9. Influence of pyramidal angle on the reflection coefficient.
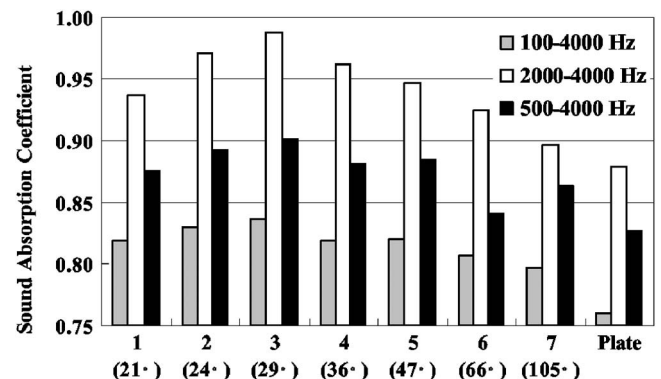
FIG. 10. Broadband averaged normal incidence sound absorption coefficients of the pyramidal polyurethane foams.

Huoy-Shyi Tsay: Normal incidence absorption of pyramidal foam

TABLE III. Broadband averaged normal incidence sound absorption coefficients of pyramidal polyurethane foams.

Model No.	1 (21°)	2 (24°)	3 (29°)	4 (36°)	5 (47°)	6 (66°)	7 (105°)	8 Plate
100–4000 Hz	0.8186	0.8300	**0.8368**	0.8190	0.8203	0.8066	0.7967	0.7602
2000–4000 Hz	0.9365	0.9706	**0.9875**	0.9624	0.9471	0.9246	0.8966	0.8787
500–4000 Hz	0.8755	0.8920	**0.9007**	0.8814	0.8847	0.8414	0.8634	0.8268

frequency ranges of 100–4k, 500–4k, and 2k–4k Hz. It is found that in all interested frequency ranges, the broadband averaged normal incidence sound absorption coefficients of the third model of the pyramidal foam are higher than that of other models. In the high frequency range (2k–4k Hz), the averaged normal incidence sound absorption coefficient of the third model of the pyramidal foam even reaches the value of 0.98. It is concluded that the third model of the pyramidal foam has the maximum broadband averaged normal incidence sound absorption coefficient among the models analyzed. Through the pyramidal foam analysis, one learned that the change of the surface shape of the foam can reach in different aspect of acoustic enhancement, and the acoustic results of the foam with arbitrary shape can be efficiently analyzed by the 3D FEFDAA.

V. CONCLUSIONS

The 3D FEFDAA, based on the Biot's dynamic poroelasticity theory, for foams with tetrahedral porous elements was introduced in this study. In addition, the use of the analysis for predictions of acoustic properties of pyramidal foams was demonstrated. The normal incidence sound absorption analyses performed in this study lead to the following conclusions.

The analysis developed in this study is more efficient than the earlier studies by saving the calculation time on the elements of the incident acoustic field, and it can be applied to design and predict the acoustic properties of three-dimensional foams such as the pyramidal foams examined in this study. The 3D FEFDAA can accurately predict the acoustic properties of the foams with suitable boundary conditions applied.

In the normal incidence sound absorption analyses of the pyramidal foams, the shape of the pyramid base and the foam volume are kept the same throughout the study. The influence of the change of the pyramidal shape is performed by changing the angle between two sides of a pyramid. It is concluded that the change of the upper surface of the foam from plane to pyramid can greatly enhance the normal incidence sound absorption capability in the mid and high frequency ranges (1k–2k and 2k–4k Hz). By reducing the pyramidal angle, the highest broadband averaged normal incidence sound absorption capability of the pyramidal foam can be pursued. It is found that the pyramidal polyurethane foam with the pyramidal angle of 29° has the highest broadband averaged normal incidence sound absorption coefficients in 100–4k, 500–4k, as well as 2k-4k Hz ranges Through this study, one learned that the change of the surface shape of the foam could reach in different aspect of acoustic enhancement.

APPENDIX A: SHAPE FUNCTION MATRIX OF THE TETRAHEDRAL POROUS ELEMENT

$$
M = \begin{bmatrix} [M]_1 \\ [M]_2 \\ [M]_3 \\ [M]_4 \\ [M]_5 \\ [M]_6 \end{bmatrix} = \begin{bmatrix} M_1 & 0 & M_2 & 0 & M_3 & 0 & M_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \cdots & & \cdots & & \cdots & & \cdots & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & M_1 & 0 & M_2 & 0 & M_3 & 0 & M_4 \end{bmatrix}. \tag{A1}
$$

APPENDIX B: DYNAMIC EQUATION OF THE TETRAHEDRAL POROUS ELEMENT

$$\sum_{e=1}^{n} \int_{V_e} [[M]_1^T, [M]_2^T, [M]_3^T, [M]_4^T, [M]_5^T, [M]_6^T] \begin{Bmatrix} s^2(\rho_{11}\bar{u}_x + \rho_{12}\bar{U}_x) + bs(\bar{u}_x - \bar{U}_x) \\ s^2(\rho_{11}\bar{u}_y + \rho_{12}\bar{U}_y) + bs(\bar{u}_y - \bar{U}_y) \\ s^2(\rho_{11}\bar{u}_z + \rho_{12}\bar{U}_z) + bs(\bar{u}_z - \bar{U}_z) \\ s^2(\rho_{12}\bar{u}_x + \rho_{22}\bar{U}_x) - bs(\bar{u}_x - \bar{U}_x) \\ s^2(\rho_{12}\bar{u}_y + \rho_{22}\bar{U}_y) - bs(\bar{u}_y - \bar{U}_y) \\ s^2(\rho_{12}\bar{u}_z + \rho_{22}\bar{U}_z) - bs(\bar{u}_z - \bar{U}_z) \end{Bmatrix} dV$$

$$-\sum_{e=1}^{n} \int_{l_e} [[M]_1^T, [M]_2^T, [M]_3^T, [M]_4^T, [M]_5^T, [M]_6^T] \begin{Bmatrix} \bar{\sigma}_{xx}l_x + \bar{\tau}_{xy}l_y + \bar{\tau}_{xz}l_z \\ \bar{\tau}_{xy}l_x + \bar{\sigma}_{yy}l_y + \bar{\tau}_{yz}l_z \\ \bar{\tau}_{xz}l_x + \bar{\tau}_{yz}l_x + \bar{\sigma}_{zz}l_z \\ \bar{\sigma}l_x \\ \bar{\sigma}l_y \\ \bar{\sigma}l_z \end{Bmatrix} dl + \sum_{e=1}^{n} \int_{V_e} [k]^T [\hat{q}] dV = 0, \tag{B1}$$

where $\bar{\sigma}_{xx}$, $\bar{\sigma}_{yy}$, $\bar{\sigma}_{zz}$, $\bar{\tau}_{xy}$, $\bar{\tau}_{xz}$, $\bar{\tau}_{yz}$, and $\bar{\sigma}$ are the Laplace transforms of σ_{xx}, σ_{yy}, σ_{zz}, τ_{xy}, τ_{xz}, τ_{yz}, and σ, respectively. σ_{xx}, σ_{yy}, σ_{zz}, τ_{xy}, τ_{xz}, and τ_{yz} are stresses acting on the solid skeleton of the foam and σ is the stress of the air in the foam

$$[\hat{q}] = \left\{ \frac{\partial \bar{u}_x}{\partial x}, \frac{\partial \bar{u}_x}{\partial y}, \frac{\partial \bar{u}_x}{\partial z}, \frac{\partial \bar{u}_y}{\partial x}, \frac{\partial \bar{u}_y}{\partial y}, \frac{\partial \bar{u}_y}{\partial z}, \frac{\partial \bar{u}_z}{\partial x}, \frac{\partial \bar{u}_z}{\partial y}, \frac{\partial \bar{u}_z}{\partial z}, \frac{\partial \bar{U}_x}{\partial x}, \frac{\partial \bar{U}_x}{\partial y}, \frac{\partial \bar{U}_x}{\partial z}, \right.$$
$$\left. \frac{\partial \bar{U}_y}{\partial x}, \frac{\partial \bar{U}_y}{\partial y}, \frac{\partial \bar{U}_y}{\partial z}, \frac{\partial \bar{U}_z}{\partial x}, \frac{\partial \bar{U}_z}{\partial y}, \frac{\partial \bar{U}_z}{\partial z} \right\}^T,$$

$$[k] = [k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10}, 0, 0, 0, k_{11}, 0, 0, 0, k_{12}],$$

where

$$k_1 = (A + 2N)\frac{\partial [M]_1^T}{\partial x} + A\left(\frac{\partial [M]_2^T}{\partial y} + \frac{\partial [M]_3^T}{\partial z}\right) + Q\left(\frac{\partial [M]_4^T}{\partial x} + \frac{\partial [M]_5^T}{\partial y} + \frac{\partial [M]_6^T}{\partial z}\right),$$

$$k_2 = k_4 = \left(\frac{\partial [M]_1^T}{\partial y} + \frac{\partial [M]_2^T}{\partial x}\right), \quad k_3 = k_7 = \left(\frac{\partial [M]_1^T}{\partial z} + \frac{\partial [M]_3^T}{\partial x}\right),$$

$$k_5 = (A + 2N)\frac{\partial [M]_2^T}{\partial y} + A\left(\frac{\partial [M]_1^T}{\partial x} + \frac{\partial [M]_3^T}{\partial z}\right) + Q\left(\frac{\partial [M]_4^T}{\partial x} + \frac{\partial [M]_5^T}{\partial y} + \frac{\partial [M]_6^T}{\partial z}\right),$$

$$k_6 = k_8 = \left(\frac{\partial [M]_2^T}{\partial z} + \frac{\partial [M]_3^T}{\partial y}\right),$$

$$k_9 = (A + 2N)\frac{\partial [M]_3^T}{\partial z} + A\left(\frac{\partial [M]_1^T}{\partial x} + \frac{\partial [M]_2^T}{\partial y}\right) + Q\left(\frac{\partial [M]_4^T}{\partial x} + \frac{\partial [M]_5^T}{\partial y} + \frac{\partial [M]_6^T}{\partial z}\right),$$

$$k_{10} = k_{11} = k_{12} = Q\left(\frac{\partial [M]_1^T}{\partial x} + \frac{\partial [M]_2^T}{\partial y} + \frac{\partial [M]_3^T}{\partial z}\right) + R\left(\frac{\partial [M]_4^T}{\partial x} + \frac{\partial [M]_5^T}{\partial y} + \frac{\partial [M]_6^T}{\partial z}\right).$$

[1] N. Voronina, "An empirical model for rigid frame porous materials with high porosity," Appl. Acoust. **51**(2), 181–198 (1997).

[2] D. Lafarge, P. Lemarinier, J. F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**(4), 1995–2006 (1997).

[3] M. A. Biot, "Theory of elasticity and consolidation for a porous anisotropic solid," J. Appl. Phys. **26**(2), 182–185 (1955).

[4] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid I. Low-frequency range, II: Higher frequency range," J. Acoust. Soc. Am. **28**(2), 168–191 (1956).

[5] M. A. Biot and D. G. Willis, "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601 (1957).

[6] A. Craggs, "A finite element model for rigid porous absorbing materials," J. Sound Vib. **61**(1), 101–111 (1978).

2692   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Huoy-Shyi Tsay: Normal incidence absorption of pyramidal foam

[7] A. Craggs, "Coupling of finite element acoustic absorption models," J. Sound Vib. **66**(4), 605–613 (1979).

[8] V. Easwaran and M. L. Munjal, "Finite element analysis of wedges used in anechoic chambers," J. Sound Vib. **160**(2), 333–350 (1993).

[9] W.-H. Chen, F.-C. Lee, and D.-M. Chiang, "On the acoustic absorption of porous materials with different surface shapes and perforated plates," J. Sound Vib. **237**(2), 337–355 (2000).

[10] Y. J. Kang and J. S. Bolton, "Optimal design of acoustical foam treatments," J. Vibr. Acoust. **118**, 498–504 (1996).

[11] H.-S. Tsay and F.-H. Yeh, "Frequency response function for prediction of planar cellular plastic foam acoustic behavior," J. Cell. Plast. **41**(2), 101–131 (2005).

[12] H.-S. Tsay and F.-H. Yeh, "Finite element frequency-domain acoustic analysis of open-cell plastic foams," Finite Elem. Anal. Design **42**(4), 314–339 (2006).

[13] H.-S. Tsay and F.-H. Yeh, "The influence of circumferential edge constraint on the acoustical properties of open-cell polyurethane foam samples," J. Acoust. Soc. Am. **119**(5), 2804–2814 (2006).

M-ary frequency shift keying communications over an underwater acoustic channel: Performance comparison of data with models

Wen-Bin Yang and T. C. Yang
Naval Research Laboratory, Washington DC 20375

The bit error rate (BER) is modeled for *M*-ary frequency-shift-keying (MFSK) signals in a fading channel using different envelope amplitude statistics and compared with frequency-hopping (FH) FSK signaling data collected during the TREX04 experiment, which used Binary FSK (BFSK) for each symbol. The BER data is found to be significantly higher than that of a Rayleigh fading channel and is found to agree with that modeled with a *K*-distribution amplitude fading distribution. Using the parameters of the *K*-distribution, independently measured from the *M*-sequence data collected in the same experiment, the modeled BER as a function of the signal-to-noise ratio (SNR) agrees well with the measured BER within the SNR of the data. This suggests that the BER of FH-FSK signaling can be predicted given some prior channel characterization measurements. [DOI: 10.1121/1.2354463]

I. INTRODUCTION

Underwater acoustic communications are different from radio frequency (RF) communications in many aspects. Most notable are the limited bandwidth available for communications to a distance due to frequency-dependent propagation loss, the long multipath delays that cause intersymbol interference, and the rapid signal fluctuations due to random ocean media. The differences become an important element in the receiver design in order to optimize the communication performance. Likewise, performance modeling and the ability to predict performance depend on knowledge of the channel properties, such as the signal envelope amplitude distributions. Performance predictions for RF communications are generally not applicable to underwater acoustic channels without modifications.

Bit error rate (BER) performance in RF communications has been successfully modeled using the Rayleigh or Rician envelope amplitude distribution.[1–3] In this paper we analyze the BER for frequency-hopping frequency-shift-keying (FH-FSK) signals based on at-sea data collected during the TREX04 experiment. It is found that the BER for underwater acoustic channels is higher than that of a rf channel given the same signal-to-noise (SNR) ratio. The reason is that the envelope amplitude statistics for FSK signals do not follow the Rayleigh or Rician envelope amplitude distribution.

The FSK signaling scheme is referred to as incoherent communications, since detection of the signal symbols is based on the symbol energy only; no phase information is required. It is thus less sensitive to the channel fluctuations and provides a more reliable communication link. FSK signaling is widely used in RF communications due to the (relatively) short multipath delay and the wide bandwidth available. For multipath underwater acoustic communications, the FSK signaling scheme has a disadvantage in that it yields a low data rate; the longer the multipath delay, the slower the rate, since the symbol duration is generally on the order of the multipath spread. Despite this shortcoming, the FSK signal is useful for many applications. One of them is command and control of a remotely operable vehicle, which can tolerate a low data rate but requires a highly reliable acoustic communication link. Note that FH-FSK signaling, using the spread spectrum method, allows multiple users to communicate at the same time (multiaccess communications) with minimum interference between the users, without requiring (precise) time synchronization between the signals. The requirement for multiaccess communications between mobile platforms with variable communication latency provides the motivation to conduct the BER analysis for FH-FSK communications.

For underwater acoustic communications, the rapid medium fluctuations and the long multipath delays of the acoustic signals significantly increase the BER compared with that found in RF communications. How much higher is the BER, and for that matter, how to predict the BER for FSK signals for underwater channels, is the focus of this paper. To calculate the BER for FSK signals, one needs a model of the envelope amplitude probability distribution that fits the data. Previous data analysis by the authors has shown that the envelope amplitude distribution of *M*-ary FSK (MFSK) follows the *K*-distribution.[4] Hence, the *K*-distribution will be included in this paper, along with the Rayleigh and Rician distributions as candidate models for BER predictions. The purpose is to find the model that best fits with the BER data. For this study, FH-FSK data for a single user will be used to avoid interferences caused by multiusers. In this case, FH-FSK signals are treated equivalently as FSK signals (without hopping), assuming that the probability distributions for different frequency components are independent and identically distributed. BER for multiuser communications is beyond the scope of this paper.

II. ENVELOPE AMPLITUDE STATISTICS AND BER MODELING

BER for FSK communications has been modeled (for RF communications) using Rayleigh and Rician distributions. They are two of the models used here for comparisons. In addition, we also calculate the BER using the K-distribution.

An introduction of the K-distribution is in order here. The K-distribution has been used to model monostatic reverberation returns for high-resolution sonar.[5,6] As the sonar time resolution improves, the insonified area for backscattering decreases and the scattered return becomes less Rayleigh-like. The K-distribution is one of the non-Rayleigh distributions, which are characterized by a high-tail distribution; namely, the percentage of its high amplitude distribution is higher than the Rayleigh distribution.

For high frequency sound propagation through a random ocean, recent measurements of the envelope amplitude distribution based on experimental data (M-sequence) have suggested that it is non-Rayleigh like and behaves like the K-distribution.[4] The reason is that the envelope amplitudes of high frequency signals follow a joint probability distribution described by a short-term (short-time scale) probability distribution conditioned on a long-term (long-time scale) probability distribution.[4] The short term fluctuation is presumably caused by the microfine structures in the ocean medium (e.g., turbulence) that have a time scale of seconds or less, whereas the long term fluctuation is caused by the fine structures (e.g., internal waves) that have a time scale of minutes or more. The parameters of the K-distributions were measured using a pseudorandom (M-sequence) signal.[4]

An empirical approach is taken in this paper using three probability distribution functions to fit data. It will be shown that (i) the BER data for FH-FSK signals are better fitted with the K-distribution model than either the Rayleigh or Rician model, and (ii) the parameters of the K-distribution determined from the FH-FSK signals agree well with that deduced previously from the M-sequence data. This implies that one can potentially predict the FH-FSK BER given some prior channel characterization measurements.

In this section, the Rician, Rayleigh, and K-distribution will be briefly reviewed and used to calculate the BER for MFSK signals. The BER predictions are compared with data in the next section.

A. Rician/Rayleigh distribution

Rayleigh and Rician distributions have been used to model a mobile radio channel in which the (narrow band) signal amplitudes (in-phase and quadrature components) are assumed Gaussian distributed. For the Rician distribution, the probability density function for the envelope amplitude z is given by[7]

$$p_z(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2 + \mu^2}{2\sigma^2}\right) I_0\left(\frac{\mu z}{\sigma^2}\right), \quad z \geq 0, \tag{1}$$

where z is the amplitude; μ and σ depend upon means and variances of the in-phase and quadrature components, and I_0 is the modified Bessel function of the first kind of zeroth

order. The first and second moments of Z are[7]

$$E\{Z\} = \sqrt{\frac{\pi}{2}} \sigma \cdot \exp(-\overline{K}/2) \cdot [(1 + \overline{K}) \cdot I_0(\overline{K}/2)$$
$$+ \overline{K} \cdot I_1(\overline{K}/2)], \tag{2}$$

$$E\{Z^2\} = 2\sigma^2 + \mu^2 = 2\sigma^2(1 + \overline{K}), \tag{3}$$

where $\overline{K} = \mu^2/2\sigma^2$ is a Rician parameter and is defined as the power ratio of the deterministic (specular) to random (scatter) component of the signal. I_1 is the modified Bessel function of the first kind of first order.

The Rayleigh distribution is a special case of the Rician distribution when $\mu = 0$ (or $\overline{K} = 0$). The Rayleigh distribution is as below.

$$p_z(z) = \frac{z}{\sigma^2} \exp\left(-\frac{z^2}{2\sigma^2}\right), \quad z \geq 0. \tag{4}$$

The first and second moments of the Rayleigh distribution are given by

$$E(Z) = \sqrt{\pi/2} \cdot \sigma \tag{5}$$

and

$$E(Z^2) = 2\sigma^2. \tag{6}$$

B. *K*-distribution

The K-distribution (see the Appendix) is parametrized with a shape parameter v and scale parameter α[5,6]

$$p_z(z) = \frac{4}{\sqrt{\alpha}\Gamma(v)} \left(\frac{z}{\sqrt{\alpha}}\right)^v K_{v-1}\left(\frac{2z}{\sqrt{\alpha}}\right), \quad z \geq 0, \quad \alpha > 0,$$
$$v > 0, \tag{7}$$

where K_{v-1} is the modified Bessel function of the second kind, of order $v-1$, and $\Gamma(v)$ is the Gamma function. The first and second moments of the K-distribution are as below.

$$E(Z) = \frac{\sqrt{\alpha\pi}}{2} \cdot \frac{\Gamma(v + 0.5)}{\Gamma(v)} \tag{8}$$

and

$$E(Z^2) = \alpha v. \tag{9}$$

A special case of the K-distribution is a Rayleigh distribution, which is obtained as $v \to \infty$ with αv remaining as a constant, $\alpha v = 2\sigma^2$.

To compare the characteristics of the various distributions, the amplitude probability distributions for Rayleigh, Rician, and K-distributions are plotted in Fig. 1(a), under the condition that they have the same second moment (the same average signal intensity), i.e., $E(Z^2) = 1$. For the Rayleigh distribution, one has $\sigma = 1/\sqrt{2}$. For the Rician distribution, $\sigma = 1/\sqrt{2(1 + \overline{K})}$, where $\overline{K} = 1$, 3.16, and 10 (corresponding to 0, 5, and 10 dB) are used. For the K-distribution one has $\alpha = 1/v$ with $v = 1$, 2, and 5.

One notes that the mean amplitude of the distribution is the largest for the Rician distribution, followed by the Rayleigh distribution, and the K-distribution. In terms of the
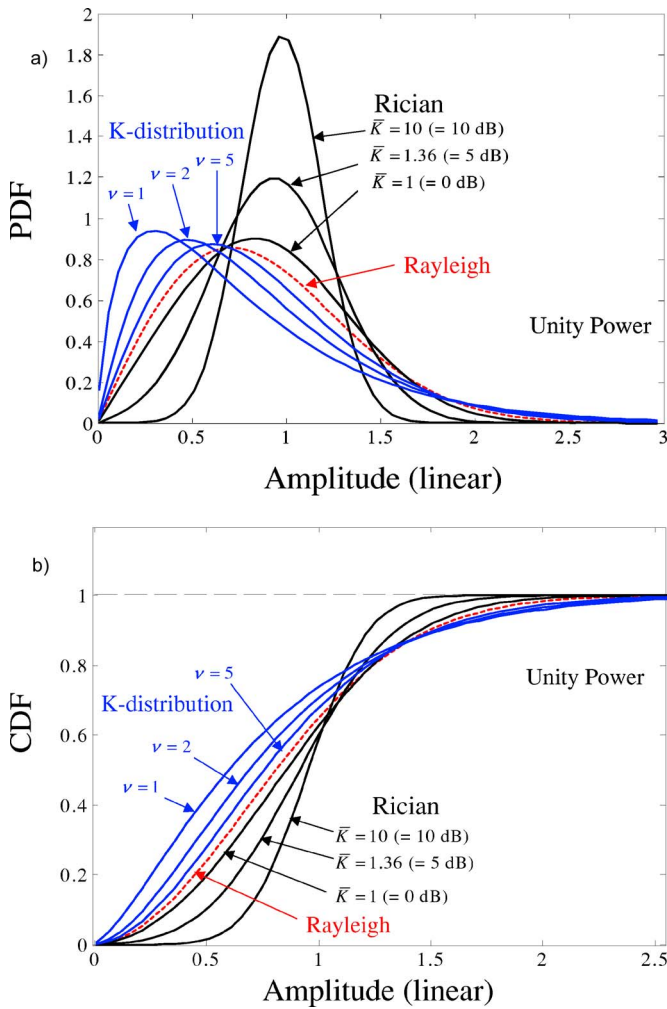
FIG. 1. (Color online) A comparison of Rician/Rayleigh and K-distribution statistics in terms of (a) probability distribution functions and (b) cumulative (probability) density functions.

physical model of the signal, the Rayleigh distribution is described by randomly scattered arrivals. The Rician model describes a signal with a deterministic component plus a random component; hence, the peak distribution has higher amplitudes than the Rayleigh model. The K-distribution has a long tail for the high amplitude distribution. Relative to the high amplitude tail, the "random" component is concentrated at low amplitudes. The different nature of the distributions can be quantitatively summarized using the cumulative density functions (CDF), which are plotted in Fig. 1(b) for the three distributions. One notes also that the K-distribution has the highest CDF for low percentile (e.g., $<30\%$) distributions, followed by the Rayleigh and Rician distribution. For the high percentile (e.g., $>85\%$) distributions, the Rician distribution has the highest CDF, followed by the Rayleigh and K-distributions. These differences in the probability distributions have a direct bearing on the BER, as described later.

C. BER via an energy detector

In this section, the BER for a given SNR for MFSK signals is calculated using an energy detector, which is an incoherent receiver. For the incoherent MFSK, the conditional probability of symbol error under a given received signal amplitude z is[8]

$$p_s(\varepsilon|z) = \frac{1}{M}\sum_{i=2}^{M}(-1)^i\binom{M}{i}\exp\left[-\left(1-\frac{1}{i}\right)\rho\right], \quad (10)$$

where $M=2^q$ (q bits per symbol), σ_N^2 is the variance of Gaussian noise, and $\rho=z^2/2\sigma_N^2$ is the SNR.

Now consider the fact that the signal has a fading distribution. The probability of symbol error is then given by

$$p_s(\varepsilon) = \int_0^\infty p_s(\varepsilon|z)\cdot p_z(z)\cdot dz. \quad (11)$$

The relationship between the probability of symbol error and the probability of bit error for the MFSK signal is given by[3]

$$p_b(\varepsilon) = \frac{M}{2(M-1)}p_s(\varepsilon). \quad (12)$$

The probability of bit error for MFSK over a Rician fading channel can be derived using Eq. (1) and Eqs. (10)–(12). (For a derivation, see, for example, Ref. 9.) The probability of bit error is found to be

$$p_b(\varepsilon) = \frac{1}{2(M-1)}\sum_{i=2}^{M}(-1)^i\binom{M}{i}\frac{\dfrac{1+\overline{K}}{m_\rho}}{\dfrac{1+\overline{K}}{m_\rho}+1-\dfrac{1}{i}}$$

$$\times\exp\left(-\frac{\overline{K}\left(1-\dfrac{1}{i}\right)}{\dfrac{1+\overline{K}}{m_\rho}+1-\dfrac{1}{i}}\right), \quad (13)$$

where $\overline{K}$ is the Rician parameter defined above, and m_ρ denotes the mean signal power over the noise power (mean SNR), which for the Rician distribution is: $m_\rho=E(z^2)/2\sigma_N^2 = \sigma^2(1+\overline{K})/\sigma_N^2$. The mean SNR is related to the SNR per bit by $m_\rho=qE_b/N_0$, where E_b is signal energy per bit and N_0 is power spectrum density of Gaussian noise.[3]

The probability of bit error for the BFSK signal assuming a Rayleigh distribution is obtained by setting $\overline{K}=0$ and $M=2$ in the above equation. One finds

$$p_b(\varepsilon) = (2+m_\rho)^{-1} = (2+E_b/N_0)^{-1}. \quad (14)$$

Assuming, on the other hand, that the amplitude distribution is given by the K-distribution, the probability of bit error for a MFSK signal can be derived using Eq. (7) and Eqs. (10)–(12). The derivation is given in the appendix. One finds

$$p_b(\varepsilon) = \frac{1}{2(M-1)}\sum_{i=2}^{M}(-1)^i\binom{M}{i}\frac{\nu}{\Gamma(\nu)}$$

$$\times\int_0^\infty u^{\nu-1}e^{-u}\frac{1}{\nu+\left(1-\dfrac{1}{i}\right)qu\dfrac{E_b}{N_0}}du. \quad (15)$$

For BPSK, the probability of bit error reduces to

W. Yang and T. C. Yang: Underwater acoustic communications channels
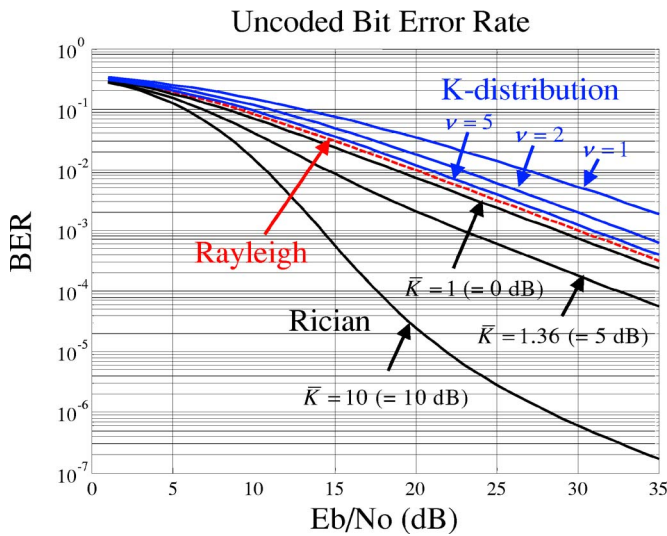
FIG. 2. (Color online) Uncoded bit error rate as a function of the E_b/N_0 for Rayleigh, Rician, and K-distribution fading channels.

$$p_b(\varepsilon) = \frac{v}{\Gamma(v)} \int_0^\infty u^{v-1} e^{-u} \left(2v + u\frac{E_b}{N_0}\right)^{-1} du. \quad (16)$$

The above probability of bit error is evaluated numerically to determine the BER for the BPSK signal. Figure 2 shows the BER as a function of the SNR per bit E_b/N_0, for the Rayleigh, Rician, and K-distributions, with the parameters given above. One notes that the Rician distribution with $\overline{K}=10$ dB has the lowest BER, whereas the K-distribution with $v=1$ has the highest BER. The BER for the Rayleigh distribution lies in between. The reason is that bit errors are normally caused by weak signals (low SNR). As shown in Fig. 1(b), the K-distribution has a higher CDF than the Rayleigh distribution for low percentile amplitudes; hence, it has a higher bit error than the Rayleigh distribution for the same SNR. Likewise, the Rician distribution has a lower CDF than

the Rayleigh distribution for low percentile amplitudes, and hence it has a lower BER than the Rayleigh distribution for the same SNR. It is noted that BER is orders of magnitude different for the three distributions when the SNR increases beyond 10 dB. The difference becomes small when SNR goes below 10 dB, but then the BER is too large to be useful. For BER modeling and prediction, knowledge of the probability distribution will be required.

III. EXPERIMENTAL DATA ANALYSIS AND COMPARISON WITH MODEL

Underwater acoustic communication data were collected during TREX04 experiment, which took place in April 2004 on the continental shelf of New Jersey coast water. Water depth in the experimental area is about 70 meters. The sound speed can be approximated by a constant: 1473 m/s, except near the surface and the bottom, where it increases to 1480–1484 m/s. A vertical line array (VLA) was deployed at a depth of approximately 35 m. It had an aperture of ~2 m and contained 8 hydrophones with uneven spacing. A fixed source was deployed at a depth of ~35 m at a range of 3.4 km. A towed source was deployed at a depth of ~23 m. The source track for the data analyzed below is shown in Fig. 3(a). FH-FSK signals were transmitted from the fixed and towed source to the VLA at different times. The signal uses a bandwidth of 4 kHz. Because of the sound speed profile, the signal travels mainly in the duct.

The bit error rates of the FH-FSK signals are analyzed in this section using both the fixed and moving source data. FH-FSK signal have been proposed for underwater acoustic communications.[10,11] Instead of modulating/occupying all frequency bins with bits at all time, the FH-FSK signal occupies only a small group of frequency bins at one time, representing an MFSK symbol, and uses a different frequency group at the next symbol time to represent the next MFSK symbol. Normally, a prescribed sequence of the fre-
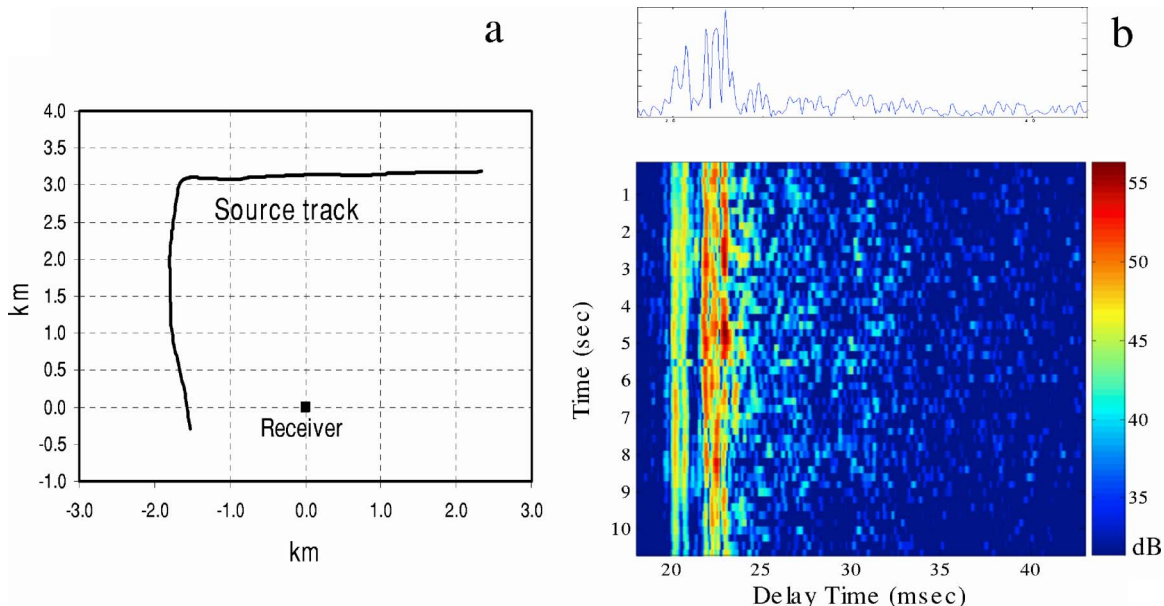
FIG. 3. (Color online) (a) Track of the moving source relative to the receiver. (b) Channel impulse response measured using M-sequence data from a fixed source to a fixed receiver.

quency bins as a function of time is used and is referred to as the hopping pattern. The pattern is repeated once it reaches the end. This modulation scheme can support multiple users, each with its own hopping pattern. An FH-FSK signal is described mathematically as

$$s(t) = \sum_i s_i(t) = \sum_i \sqrt{2P} b_i \cos[2\pi f_i(t)t] \cdot W(t - iT), \quad (17)$$

where b_i is the bit sequence as a function of symbol number i, $f_i(t)$ is the hopping frequency as a function of time, occupied by the ith symbol, P is the transmitted power at the ith symbol, T is a symbol period, and $W(t)$ is a rectangular window function representing the symbol duration

$$W(t) = \begin{cases} 1, & if \ 0 \le t < T, \\ 0, & elsewhere. \end{cases} \quad (18)$$

In the TREX04 experiment, the FH-FSK signals used covered a frequency range from 15 to 19 kHz. The hopping pattern uses 13 pairs of frequency bins, one at a time. A pair of frequencies is used to carry bit 1 or 0 (i.e., BFSK modulation for each symbol). The bin size is 80 Hz, but the frequency separation between adjacent bins is 160 Hz to minimize interfrequency interference; the symbol period is $1/80$ s. (This signaling scheme is similar to that used in the WHOI micromodem.) Each data packet transmitted contained 1680 bits lasting about 21.0 s.

As described above, FH-FSK signals were transmitted from a fixed source to the VLA, at a range of 3.4 km, with a received SNR of 26–35 dB. Figure 3(b) shows the channel impulse response function as a function of geotime, measured from the M-sequence data collected during the same experiment. It shows two dominant groups of multipath arrivals separated by $\sim$5 ms time delay, followed by groups of multipath arrivals with weaker amplitudes; the last group of arrivals is $\sim$10 dB down from the first two groups of arrivals. The total multipath spread lasts over $\sim$15 ms. One notes that each symbol, with a duration of $1/80=12.5$ ms, captures most of the multipath energy.

The data were processed with a BFSK energy detector, which compares the energy between a pair of frequencies following the hopping pattern. The received bit is declared 1 or 0 depending on which frequency among the pair has a higher energy.

We have also analyzed FH-FSK data from a towed source to the VLA. The source-receiver range covered 1.73 to 3.70 km. The SNR at the receiver was on the order of 20–27 dB. For the FH-FSK data from a moving source, the data need to be corrected for the Doppler shift first. For a towing speed of 4 knots, the maximum Doppler shift for 17 kHz is around 23 Hz. Since the frequency bin has a width of 80 Hz, about one third of the symbol energy could be shifted outside the frequency bin. One can determine the Doppler shift based on the frequency shift of a pilot tone or the time separation between a pair of pulses measuring the signal dilatation/compression. The latter method is used here. A pair of 0.1 s linear frequency modulated (LFM) signals was placed before and after the FH-FSK signal, with a time separation of approximately 26.9 s between them. The measured time separation between the LFM signals divided by the time
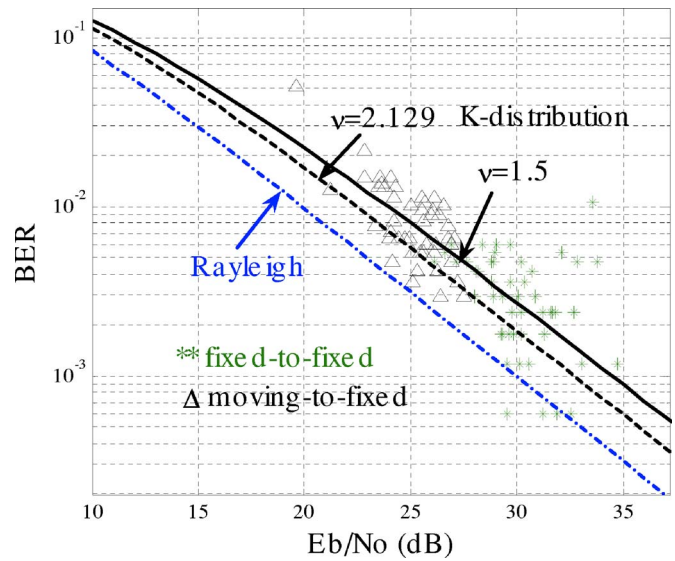
FIG. 4. (Color online) Uncoded bit error rate measured from an hour of data from a fixed-source to a fixed-receiver, and another hour of data from a moving-source to a fixed receiver.

separation of the transmitted LFM signals yields the Doppler ratio. The Doppler shift resolution is inversely proportional to the time separation of the LFM signals; it is on the order $\sim$30 mHz. The data are down- or upshifted using the Doppler shift determined from the Doppler ratio.

To determine the channel effect on the BER, one is interested in the uncoded BER, i.e., the BER in the raw data without using error correction. Figure 4 shows the measured BER as a function of the SNR (E_b/N_0) for both the fixed-fixed transmission and moving-fixed transmission data. Only one receiver is used for this analysis; no significant difference is observed between data on different receivers. Also shown in Fig. 4 is the BER modeled using the Rayleigh distribution and K-distribution. One observes in Fig. 4 that the measured BER is, on average, significantly higher than that based on the Rayleigh fading statistics model. (Consequently, there is no need to include Rician model, which has a lower BER.) It is in fair agreement with the BER calculated using the amplitude fading statistics based on the K-distribution. One finds that the data can be fitted with the K-distribution model using $\nu=1.5$. [Note that the wide spread of the BER for high SNR data is a result of a limited number of data points. The BER at high SNR is usually small. A range of 1–8 bits in error (of 1680) can occur for a small number of data samples. It corresponds to 0.1%–0.5% bit error rate, which shows up as a wide spread in a log plot.]

In a separate study, it was observed that the envelope amplitude distributions for narrowband signals of frequency bandwidth of 80 and 320 Hz follow the K-distribution with $\nu=2.129$ and 1.28, respectively.[4] The envelope amplitude distributions were based on narrow frequency bin components averaged over pseudorandom (M-sequences) signals spread over a 4 kHz bandwidth. The M-sequence data were collected during the same experiment at a slightly different time. The BER curve for $\nu=2.129$ is also shown in Fig. 4. Both values of ν are consistent with the data given the large spread of the data points. One finds that the two different

W. Yang and T. C. Yang: Underwater acoustic communications channels
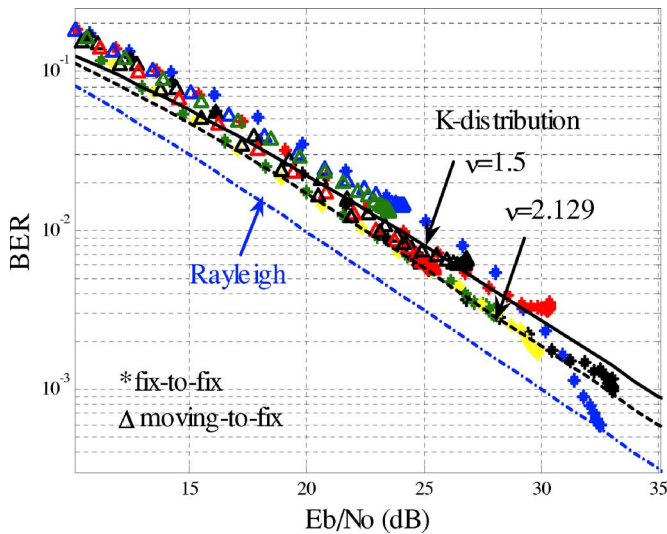
FIG. 5. (Color online) Uncoded bit error rates evaluated in terms of E_b/N_0 using 10 collected signal packets and 25 different noise data. The E_b/N_0 values are varying by different weighting of the signal and noise data.

FIG. 6. (Color online) A prediction of the uncoded bit error rate for MFSK signals, $M=2$, 4, and 8 based on K-distribution fading statistics measured in TREX04. The uncoded bit error rates are compared with those of a Rayleigh fading channel.

signals (M-sequence and FH-FSK), analyzed with two independent methods, lead to a consistent conclusion about the K-distribution.

The above result can be stated in another way. The BER for FH-FSK signals can be predicted using the measured envelope amplitude distribution function or the K-distribution with v measured from the envelope amplitude fading statistics. The prediction is found in agreement with the measured BER. Although only the frequency-hopped BFSK signals were transmitted during the TREX04 experiment, armed with the above theory, one could predict the BER for (frequency-hopped) QFSK or 8FSK signals for the TREX04 environment. The predictions and practical implications will be covered in the next section.

Figure 4 shows the BER predictions for FH-FSK signals as a function of (decreasing) SNR. To evaluate how good the prediction is against data, noise collected at sea are added to the signal data (ten different packets) to determine the BER for smaller SNR cases. The noise data were collected during the TREX04 experiment when no signal was transmitting within the frequency band. Many (25) segments of noise data collected at a different time and different locations (during the same experiment) were used and added to the 10 packets, from which the BER was determined and then averaged over the data ensembles. The result is plotted in Fig. 5 using both fixed-source and moving-source data. One finds reasonably good agreement between the model predictions (using the K-distribution) and measured BER. Again, note that the BER data is significantly higher than that predicted by the Rayleigh amplitude statistics model.

IV. BER PREDICTIONS FOR HIGHER ORDER MFSK SIGNALS

The previous sections showed that BER can be modeled and predicted for BFSK signals. Good agreement with data was demonstrated. One notes that other types of MFSK signals, particularly QFSK ($M=4$) and 8FSK ($M=8$), are also
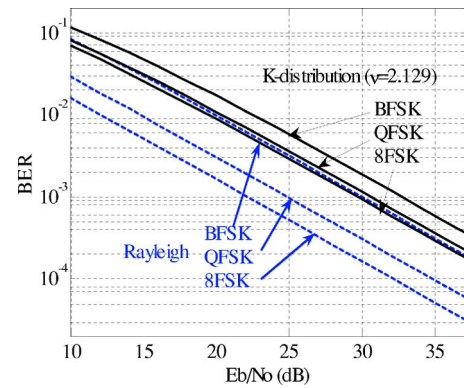
of interest; for example, Ref. 11 uses 8FSK for FH-FSK signal transmission. Unfortunately, these signals were not transmitted during the TREX04 experiment, and therefore one cannot conduct a comparison between data and models. Nonetheless, it is of interest (to the communication system designers) to know what kind BER can be expected for this type of signal under the same environment.

To analyze the BER for QFSK and 8FSK, let us assume that the frequency bin size is 80 Hz wide, as was used for BFSK. The symbol duration is $1/80$ s. For BFSK ($M=2$) signal, each symbol consists of two frequency bins. The data rate is 80 symbols/s, or 80 bits/s. It requires a bandwidth of 160 Hz (for each symbol). For the QFSK ($M=4$) signal, each symbol consists of 4 frequency bins. The data rate is 80 symbols/s or 160 bits/s. It requires a bandwidth of 320 Hz (for each symbol). The bandwidth is doubled; so is the data rate. Hence QFSK has the same bandwidth efficiency (bit rate/bandwidth) as BFSK.

For the 8FSK ($M=8$) signal, each symbol consists of 8 frequency bins. The bit rate is now $\log_2(M) \times 80$ $=240$ bits/s, but it requires a bandwidth of 640 Hz (for each symbol). Compared with BFSK, one finds that 8FSK requires eight times the bandwidth, but only increases the bit rate by three times. Continuing the analysis, one can show that for MFSK signaling, the bandwidth efficiency decreases as M increases for $M>4$. On the other hand, one expects that BER will improve as one increases M. Given the envelope amplitude distribution as determined above for underwater acoustic communications, one can calculate the improvement in BER as a function of M.

Assuming the K-distribution with its parameter (v $=2.129$) measured from the TREX 04 data, the BER is calculated using Eq. (15) for QFSK and 8FSK signals. The numerical results are plotted in Fig. 6 to compare with the BER for BFSK. Also shown is the BER using the Rayleigh distribution for comparison, which has been assumed by many system designers. The BER over a Rayleigh channel is given in Eq. (14).

From Fig. 6, one finds that BER is reduced by a factor of $\sim 2/3$ and 3.5/6 for QFSK and 8FSK, respectively, compared with BFSK. The reduction factor is almost independent of the input SNR between 10 to 35 dB. One expects that the

BER will further decrease as M increases. Hence, there is a tradeoff between BER and band efficiency as a function of M. The tradeoff is a system dependent issue, which will not be pursued here. Note, however, that between BFSK and QFSK, QFSK is a better choice since QFSK has the same bandwidth efficiency as BFSK but a lower BER (for the same E_b/N_0) than BFSK.

V. SUMMARY AND CONCLUSIONS

In this paper, the BER of FH-FSK signals were analyzed based on data collected during the TREX04 experiment. The data were transmitted from a fixed source as well as a towed source to a VLA moored to the ocean bottom. The BER data were modeled as a function of input SNR (E_b/N_0), assuming a Rayleigh and K-distribution envelope amplitude statistical distributions for a fading channel. It is found that the measured BER is significantly higher than that modeled by the Rayleigh distribution (hence is also higher than a Rician statistical model with the same E_b/N_0) and can be fitted with the K-distributions. The BER model using a K-distribution fading statistics, based on the model parameters independently estimated from the M-sequence data, yields a reasonable prediction for the BER of the FH-FSK signals. Using the same model, one can also predict the BER for QFSK and 8FSK signals under the same (TREX04) ocean environment.

ACKNOWLEDGMENTS

APPENDIX: MFSK ERROR PROBABILITY OVER A K-DISTRIBUTION FADING CHANNEL

A random variable Z, which has a K-distribution $p_z(z)$, is a product of two independent random variables, X and Y, where X follows a chi distribution with 2ν degrees of freedom and Y follows a Rayleigh distribution in terms of a power ratio α, as given next:

$$p_X(x) = \frac{2x^{2\nu-1}e^{-x^2}}{\Gamma(\nu)} \quad \text{and} \quad p_Y(y) = \frac{2ye^{-y^2/\alpha}}{\alpha}. \quad \text{(A1)}$$

Thus, the K-distribution is given by

$$p_z(z) = \int_0^\infty \frac{1}{|x|} p_Y(z/x) p_X(x) dx$$

$$= \int_0^\infty \frac{1}{x} \left(\frac{2z}{x} \frac{1}{\alpha} e^{-z^2/\alpha x^2} \right) \left(\frac{2}{\Gamma(\nu)} x^{2\nu-1} e^{-x^2} \right) dx$$

$$= \int_0^\infty \frac{4}{\alpha\Gamma(\nu)} z \cdot x^{2\nu-3} e^{-z^2/\alpha x^2} e^{-x^2} dx$$

$$= \frac{4}{\sqrt{\alpha}\Gamma(\nu)} \int_0^\infty \frac{z}{\sqrt{\alpha}} \cdot x^{2\nu-3} e^{-(z/\sqrt{\alpha}x)^2} e^{-x^2} dx$$

$$= \frac{4}{\sqrt{\alpha}\Gamma(\nu)} \left(\frac{z}{\sqrt{\alpha}} \right)^\nu K_{\nu-1}\left(\frac{2z}{\sqrt{\alpha}} \right), \quad z \geq 0, \ \alpha > 0,$$

$$\nu > 0. \quad \text{(A2)}$$

Using the above K-distribution fading statistics $p_z(z)$ and Eqs. (10) and (11), the probability of symbol error is derived as below:

$$p_s(\varepsilon) = \int_0^\infty \frac{1}{M} \sum_{i=2}^M (-1)^i \binom{M}{i} e^{-(1-1/i)z^2/2\sigma_N^2}$$

$$\times \int_0^\infty \frac{4}{\alpha\Gamma(\nu)} z \cdot x^{2\nu-3} e^{-z^2/\alpha x^2} e^{-x^2} dx \, dz$$

$$= \frac{4}{M\alpha\Gamma(\nu)} \sum_{i=2}^M (-1)^i \binom{M}{i} \int_0^\infty e^{-(1-1/i)z^2/2\sigma_N^2}$$

$$\times \int_0^\infty z \cdot x^{2\nu-3} e^{-z^2/\alpha x^2} e^{-x^2} dx \, dz$$

$$= \frac{4}{M\alpha\Gamma(\nu)} \sum_{i=2}^M (-1)^i \binom{M}{i}$$

$$\times \int_0^\infty x^{2\nu-3} e^{-x^2} \cdot \int_0^\infty z \cdot e^{-(1-1/i)z^2/2\sigma_N^2} e^{-z^2/\alpha x^2} dz \, dx$$

$$= \frac{4}{M\alpha\Gamma(\nu)} \sum_{i=2}^M (-1)^i \binom{M}{i}$$

$$\times \int_0^\infty x^{2\nu-3} e^{-x^2} \frac{\alpha\sigma_N^2 x^2}{\alpha(1-1/i)x^2 + 2\sigma_N^2} dx$$

$$= \frac{2\sigma_N^2}{M\Gamma(\nu)} \sum_{i=2}^M (-1)^i \binom{M}{i} \int_0^\infty \frac{u^{\nu-1}e^{-u}}{(1-1/i)\alpha u + 2\sigma_N^2} du, \quad \text{(A3)}$$

where $u = x^2$. In terms of the mean SNR, $m_\rho = E(Z^2)/2\sigma_N^2 = \alpha\nu/2\sigma_N^2$, the probability of symbol error becomes

$$p_s(\varepsilon) = \frac{1}{M} \sum_{i=2}^M (-1)^i \binom{M}{i} \frac{\nu}{\Gamma(\nu)}$$

$$\times \int_0^\infty u^{\nu-1} e^{-u} \frac{1}{\nu + (1-1/i)m_\rho u} du. \quad \text{(A4)}$$

Now using $M = 2^q$ (q bits per symbol) and replacing the mean SNR by the SNR per bit, $m_\rho = qE_b/N_0$, the probability of bit error can be expressed as

$$p_b(\varepsilon) = \frac{M}{2(M-1)} p_s(\varepsilon) = \frac{1}{2(M-1)} \sum_{i=2}^M (-1)^i \binom{M}{i} \frac{\nu}{\Gamma(\nu)}$$

$$\times \int_0^\infty u^{\nu-1} e^{-u} \frac{1}{\nu + \left(1 - \frac{1}{i}\right) qu \frac{E_b}{N_0}} du. \quad \text{(A5)}$$

For the BFSK ($M=2$) case, the BER is given by

$$p_b(\varepsilon) = \frac{\upsilon}{\Gamma(\upsilon)} \int_0^\infty u^{\upsilon-1} e^{-u} \left(2\upsilon + u\frac{E_b}{N_0}\right)^{-1} du. \qquad (A6)$$

As mentioned in Sec. II B, a Rayleigh distribution is a special case of the K-distribution as $\upsilon \to \infty$, with $\alpha\upsilon$ remaining as a constant. This can be shown via

$$p_b(\varepsilon) = \lim_{\upsilon \to \infty} \frac{\upsilon}{\Gamma(\upsilon)} \int_0^\infty u^{\upsilon-1} e^{-u} \left(2\upsilon + u\frac{E_b}{N_0}\right)^{-1} du$$

$$= \lim_{\upsilon \to \infty} \sum_{i=0}^\infty \frac{1}{2\Gamma(\upsilon)} \int_0^\infty u^{\upsilon+i-1} e^{-u} \left(-\frac{E_b}{2\upsilon N_0}\right)^i du$$

$$= \lim_{\upsilon \to \infty} \sum_{i=0}^\infty \frac{1}{2\Gamma(\upsilon)} \left(-\frac{E_b}{2\upsilon N_0}\right)^i \int_0^\infty u^{\upsilon+i-1} e^{-u} du$$

$$= \frac{1}{2} \sum_{i=0}^\infty \lim_{\upsilon \to \infty} \frac{\Gamma(\upsilon+i)}{\upsilon^i \Gamma(\upsilon)} \left(-\frac{E_b}{2N_0}\right)^i = \sum_{i=0}^\infty \frac{1}{2} \left(-\frac{E_b}{2N_0}\right)^i$$

$$= \frac{1}{2 + E_b/N_0}, \qquad (A7)$$

where $\lim_{\upsilon \to \infty}[\Gamma(\upsilon+i)/\upsilon^i\Gamma(\upsilon)] = 1$. Thus, the BFSK BER for a Rayleigh channel is a special case of Eq. (A6).

[1] R. J. C. Bultitude, "Measurement, characterization and modeling of indoor 800/900 Mhz radio channels for digital communications," *IEEE Communications Magazine*, June 1987, pp. 5–12.

[2] W. C. Y. Lee, *Mobile Communications Engineering* (McGraw-Hill, New York, 1993).

[3] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook* (McGraw-Hill, New York, 1994).

[4] W-B. Yang and T. C. Yang, "High frequency channel characterization for M-ary frequency shift keying underwater acoustic communications," J. Acoust. Soc. Am., **120**(5), November 2006.

[5] D. A. Abraham and A. P. Lyons, "Novel physical interpretations of K-distributed reverberation," IEEE J. Ocean. Eng. **27**, 800–813 (2002).

[6] K. D. LePage, "Statistics of broad-band bottom reverberation predictions in shallow-water waveguides," IEEE J. Ocean. Eng. **29**, 330–346 (2004).

[7] S. O. Rice, "Mathematical analysis of random noise," Bell Syst. Tech. J. **23**, 101 (1945).

[8] E. Arthurs and H. Dym, "On the optimum detection of digital signals in the presence of white Gaussion noise – A geometric interpretation of three basic data transmission systems," IRE Trans. Commun. Syst. **10**(4), 336–372 (1962).

[9] J. Sun and I. S. Reed, "Performance of MDPSK, MPSK and noncoherent MFSK in wireless Rician Fading Channels," IEEE Trans. Commun. **47**, No. 6, 813–816, June (1989).

[10] L. Freitag, M. Stojanovic, S. Singh, and M. Johnson, "Analysis of channel effects on direct-sequence and frequency-hopped spread-spectrum acoustic communication," IEEE J. Ocean. Eng. **26**, 586–593 (2001).

[11] M. D. Green and J. A. Rice, "Channel-tolerant FH-MFSK acoustic signaling for undersea communications and networks," IEEE J. Ocean. Eng. **25**, 28–39 (2000).

The experimental synthesis of random pressure fields: Methodology

Teresa Bravo[a)] and Cédric Maury
Laboratoire Roberval FRE-CNRS 2833,
Université de Technologie de Compiègne, Secteur Acoustique, Boîte Postale 20529,
F-60205 Compiègne Cedex, France

This paper presents a methodology for the off-line reproduction of random pressure fields with given spatial correlation characteristics. The simulation method is first presented, together with an outline of the signal processing techniques required. The design of an experimental setup is then detailed in relation with the nature of the simulated pressure fields. Of particular interest is the laboratory synthesis of three different types of partially correlated random excitations: an acoustic diffuse field, a grazing incident plane wave, and a turbulent boundary layer. The corresponding excitations are generated in a semianechoic chamber over a test panel using a near-field array of 16 loudspeakers driven by a set of optimal signals. The loudspeakers are positioned at a suitable distance above a sufficiently dense grid of microphones close to the simulation surface. The mutually correlated drive signals are determined from both the target correlation properties and the acoustic transfer functions measured between the loudspeakers and the microphones. This approach could provide a cost-effective method of reducing the variation of low frequency sound transmission measurements as well as simulating the propeller-induced noise and the boundary layer noise transmitted through aircraft fuselage structures. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354008]

I. INTRODUCTION

The experimental reproduction of random pressure fields in laboratory conditions is a subject of considerable interest within several areas. In building acoustics, the problem of predicting and measuring low frequency sound insulation has become more important recently due to the increasing presence of the noise sources in this frequency range.[1] The characterization of the transmission loss coefficient assumes that the incident sound field is ideally diffuse[2] whereas, in practice, test panels are often tested in a sound transmission suite with two reverberation chambers. At low frequencies, the modal behavior of the rooms influences the measurements of the sound insulation properties of test materials which can vary considerably from one laboratory to the other. This variability has been widely studied experimentally through inter-laboratory comparisons[3,4] as well as through analytical[5–7] and numerical[8,9] investigations. In order to make the measurement of the incident power more reliable at low frequencies, Bravo and Elliott[10] have proposed a novel approach based on the reconstruction of the statistical properties of an acoustic diffuse field over the surface of a test partition using a near-field array of suitably driven loudspeakers. Theoretical results have shown that this approach is efficient in providing a measure of the sound reduction index that only depends on the properties of the panel itself and, at this stage, it would benefit from experimental validations.

In the aeronautical and aerospace industries, there is a timely interest in the laboratory simulation of random loads over components subject to high fluctuating pressure levels. Acoustic progressive wave tube (APWT) facilities are currently used for simulating the dynamic response and the sonic fatigue of aeronautical structures. Adaptive acoustic controllers are developed to generate input power spectral densities (PSD) tailored to field measurements, as those induced by jet noise.[11,12] In particular, a parametric method, the pseudorandom time history simulation, has been widely used to synthesize suitable time-domain signals for the input PSDs.[13] It is based on the time-domain representation of the simulated signals as a sum of a large number of statistically independent harmonics with random phase distributions. The frequency variables of the harmonics are distributed according to a joint density function derived from the specified PSDs. Because of the central limit theorem, it is well suited for the generation of Gaussian random wave forms, but the method may not be computationally efficient for simulating random process with a large number of variables and subsequent improvements have been formulated using the fast Fourier transform.[14] In spite of this, the APWT facilities are not able to reconstruct the detailed spatial correlation patterns associated to the simulated random excitations. Such limitations were found to be critical to simulate the effect of acoustical loadings due to rotating blades over the nacelle surface of a ducted propeller airplane.[15] Under typical flow speed conditions, pressure patterns propagate in spirals at grazing incidence to the fuselage sidewall and the effect of the spatially periodic load on the structure has to be accounted for. Therefore, we found useful to investigate how a

---

[a)]Present address: iTEAM, Universidad Politècnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain. Electronic mail: tebrama@iteam.upv.es

laboratory system could synthesise propeller-induced noise excitations, readily modeled as a grazing incident plane wave of random phase and amplitude.

The aeronautical industry is also concerned with cabin noise problems and their impact on discomfort and the integrity of the electronic equipment. The noise induced by turbulent boundary layer (TBL) pressure fluctuations developed over the fuselage of well-streamlined aircraft is important during cruise conditions.[16–18] The use of design modifications and noise control measures provides some opportunity for reducing the boundary layer noise in the cabin assuming that a characterization of the TBL-induced noise can be made either from flight tests or low-noise wind tunnel experiments.

Early attempts to understand how TBL-induced responses could be measured made use of reverberant acoustic qualification testing.[19,20] However, for high subsonic Mach numbers, boundary layer turbulence is generally less efficient at exciting the individual panel resonances than a reverberant acoustic field of the same intensity.[21] Alternate approaches propose to use a set of suitably driven actuators to simulate TBL excitations, as initially proposed by Fahy in 1966,[22] or to directly reconstruct the TBL-induced velocity response of a test panel using a limited number of point forces.[23] In order to reconstruct TBL excitations, Fahy observed that it would, in theory, be feasible to generate a spatially decaying correlation function over a test panel using a set of plane waves impinging at continuous random angles within a specific incident range: an equal probability between grazing and normal incidence would reproduce an acoustic diffuse field excitation whereas a probability confined to one incidence angle would generate a progressive plane wave excitation of random amplitude. The spatial decay of the correlation would thus be controlled by the angular bandwidth of the incident waves, although achieving continuous incident range might be questionable using a finite number of acoustic sources. In practice, the use of a near-field array of loudspeakers would require the multichannel discrete simulation of a two-dimensional random sound field over a grid of microphones close to the test surface: the loudspeakers should be driven by optimal time-domain signals to ensure that the statistical properties of the microphones outputs are mostly similar to those from the target random field.

Such an approach can be performed in frequency domain prior to generating the time-domain driving signals. The method consists in factorizing, for each frequency, the specified cross spectral density (CSD) matrix as a product between two matrices which are the Hermitian transform of each other.[13] When driven by a set of uncorrelated white noise signals, these matrices can be considered as shaping filters, the output of which are the simulated spectra in frequency domain. The random wave forms are then generated by calculating the inverse discrete Fourier transform (IDFT) of the complex spectra. The shaping filters can *a priori* be derived from an eigendecomposition or spectral decomposition of the specified CSD matrices. The technique of simulation can also benefit from alternate decompositions such as the Cholesky factorization when the CSD matrix is Hermitian non-negative.[24]

Recently, the multichannel simulation of a TBL excitation over a plane grid of microphones using a near-field array of acoustic sources has been theoretically investigated.[25–27] But, to the best of our knowledge, the experimental reproduction of two-dimensional random processes such as those described in the above using an array of near-field loudspeakers has not yet been achieved. A recent preliminary experiment studied the invertibility of the plant matrix between a set of loudspeakers and microphones housed within an enclosure on top of which is mounted a test panel.[27] It was recommended that the enclosure modal resonances should be heavily damped so that the plant transfer functions can be equalized.

In order to avoid stringent multipaths environments in the area of simulation, a loudspeakers cabinet radiating in "free-field" at a variable height from the test surface is designed.[28] The array of loudspeakers is installed in a nonreactive acoustical environment, a semianechoic chamber, for the near-field generation of prescribed random pressures at a number of microphones positions over the test surface. In particular, we aim to simulate the statistical properties of three types of wall-pressure excitations: an acoustic diffuse field, a grazing plane wave, and a TBL pressure field. The present work first outlines the methodology used for the simulation as well as the signal processing background required for the synthesis experiment. Then a discussion about the design of the experimental facility is detailed in relation with the nature of the simulated pressure fields. The experimental results are presented in a companion paper[29] regarding the physical limitation performances and the practical feasibility of synthesizing these random pressure fields.

II. THEORETICAL BACKGROUND FOR RANDOM PRESSURE FIELD SYNTHESIS

A. The synthesis method

The theoretical formulation for the laboratory reproduction of spatially correlated random pressure fields has been detailed in previous studies for the synthesis of a TBL wall-pressure field[25–27] and for the characterization of low frequency sound transmission between two rooms via a flexible panel.[10] The formulation can be readily extended to the synthesis of any random pressure field with prescribed spatial correlation properties as long as the pressure excitations we aim to reproduce are modeled as random processes stationary both in space and time. The general synthesis method is now briefly summarized.

We consider an array of evenly spaced microphones situated over the panel surface. The microphone outputs provide a measure of the spatial variation of the pressure field, which ideally corresponds to the desired one, $\mathbf{d}$. One assumes that $\mathbf{d}$ can be derived from a set of uncorrelated white noise reference signals, $\mathbf{x}$, via a matrix of shaping filters, $\mathbf{D}$, such that $\mathbf{d} = \mathbf{Dx}$. The CSD matrix between the elements of the desired pressure field can be written as

$$\mathbf{S}_{dd} = E[\mathbf{dd}^H] = \mathbf{DS}_{xx}\mathbf{D}^H = \mathbf{DD}^H, \tag{1}$$

where E denotes the expectation operator, and ^{H} denotes the Hermitian, complex conjugate transpose. The diagonal terms of $\mathbf{S}_{dd}$ correspond to the PSDs of the pressures at the microphones and the off-diagonal terms correspond to the CSDs of the pressures between two microphones. An eigendecomposition of the Hermitian matrix $\mathbf{S}_{dd}$ reads

$$\mathbf{S}_{dd} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H = \mathbf{Q}\mathbf{\Lambda}^{1/2}(\mathbf{Q}\mathbf{\Lambda}^{1/2})^H, \tag{2}$$

where $\mathbf{Q}$ is the matrix of eigenvectors and $\mathbf{\Lambda}$ is the matrix of real, positive eigenvalues, whose square roots are denoted $\mathbf{\Lambda}^{1/2}$. Equation (2) is equal to Eq. (1) for the following matrix of shaping filters:

$$\mathbf{D} = \mathbf{Q}\mathbf{\Lambda}^{1/2}. \tag{3}$$

A matrix of control filters, $\mathbf{W}$, is determined that generates the optimum input signals to an array of loudspeakers, which drive the sensor outputs, $\mathbf{y}$, to be as close as possible to those due to the desired pressure field, $\mathbf{d}$. The vector of error signals is defined to be

$$\mathbf{e} = \mathbf{d} - \mathbf{y} = (\mathbf{D} - \mathbf{GW})\mathbf{x}, \tag{4}$$

where $\mathbf{G}$ is the matrix of acoustic transfer functions between the near field loudspeakers and the microphones. $\mathbf{W}$ thus minimizes the sum of the mean square error signals, $\mathrm{Tr}[\mathbf{ee}^H]$, where Tr is the trace operator. It can be shown that the optimal least-squares matrix of control filters is given by[30]

$$\mathbf{W}_{\mathrm{opt}} = [\mathbf{G}^H\mathbf{G}]^{-1}\mathbf{G}^H\mathbf{D} = \mathbf{G}^{\dagger}\mathbf{D}, \tag{5}$$

where $\mathbf{G}^{\dagger}$ is the pseudoinverse of $\mathbf{G}$. The calculation of $\mathbf{W}_{\mathrm{opt}}$ requires the knowledge of both the matrices $\mathbf{G}$ and $\mathbf{D}$.

Three models of wall-pressure excitations are now considered: an acoustic diffuse field, a grazing incident plane wave and a TBL. The CSD function between the pressures at two microphones, when subject to a diffuse incident sound field, is known to be of the form[31]

$$S_{dd}(r;\omega) = S_0(\omega)\frac{\sin kr}{kr}, \tag{6}$$

where $S_0(\omega)$ is the power spectral density of the pressure at any single microphone, $k = 2\pi/\lambda$ is the acoustic wave number, λ is the acoustic wavelength, and r is the distance between two microphones. We note that samples of a diffuse pressure field measured at microphones half a wavelength or more apart are uncorrelated.

The CSD function for an incident plane wave with random phase and amplitude is given by

$$S_{dd}(r;\omega) = S_0(\omega)e^{jk(|r_x|\sin\theta_K\sin\phi_L)}e^{jk(|r_y|\sin\theta_K\cos\phi_L)}, \tag{7}$$

where $|r_x|$ is the separation distance in the x direction, $|r_y|$ is the separation distance in the y direction, ϕ_L is the azimuthal incident wave angle, and θ_K is the corresponding polar angle. For a grazing incident plane wave, we have chosen $\theta_K = 90°$ and $\phi_L = 0°$.

The CSD between the pressures at two microphones beneath a TBL is assumed to be that of the Corcos model[32]
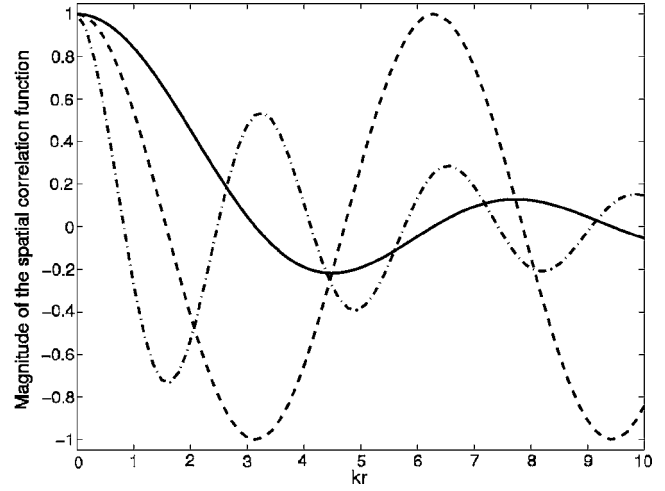
FIG. 1. Theoretical correlation as a function of the normalized separation distance between two microphones, respectively, for a diffuse pressure field (solid), a grazing incident plane wave (dashed), and a streamwise TBL pressure field (dash-dotted).

$$S_{dd}(r;\omega) = S_0(\omega)e^{-|r_x|/L_x}e^{-|r_y|/L_y}e^{-j\omega r_y/U_c}, \tag{8}$$

where L_x is the correlation length in the spanwise or x direction, L_y is the correlation length in the streamwise or y direction, and U_c is the convection velocity, which corresponds to $0.6U_\infty$, where U_∞ is the free-stream velocity. As suggested by Corcos, the correlation lengths are assumed to be inversely proportional to frequency, and have the form

$$L_x = \frac{\alpha_x U_c}{\omega}, \tag{9a}$$

$$L_y = \frac{\alpha_y U_c}{\omega}, \tag{9b}$$

where α_x and α_y are empirical constants taken to be, respectively, 1.2 and 8. The convection velocity is assumed to be 135 ms^{-1}, which corresponds to a flow Mach number of 0.66. The empirical Corcos model is well suited to describe the statistics of TBL wall-pressure fluctuations induced by high speed subsonic flows such as in aeronautical applications where the correlation lengths are large. However, it is of limited use to describe the TBL induced by low speed flows such as in marine applications where the correlation lengths are much smaller. Indeed, the short correlation scales involved are poorly resolved due to the finite size of the transducers and the attenuation resulting in the wall-pressure power spectral density should be compensated using a frequency-dependent correction factor.[32] Furthermore, the Corcos model overpredicts by 20 dB or more the spectral levels in the low wave number region. In most low Mach number applications, one may use the more comprehensive model by Chase.[33]

The spatial correlation functions of the three types of excitations are shown in Fig. 1 at a frequency of 1.6 kHz and for a variable distance to a reference microphone. It can be seen that the diffuse pressure field is correlated over an area much larger than the TBL pressure field. It will be shown in
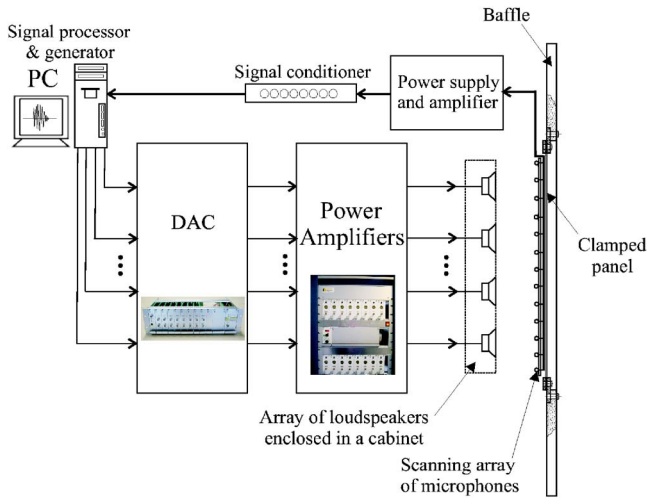
FIG. 2. Experimental setup for the laboratory synthesis of random pressure fields.

a following paper that these differences in the correlation lengths greatly influence the number of loudspeakers required in a synthesis experiment.[29]

B. Signal processing background for the synthesis experiment

The hardware implementation of the loudspeakers array synthesis experiment is first presented, together with an outline of the signal processing techniques required to generate off-line the optimal driving signals.

1. Hardware description

The experimental scheme shown in Fig. 2 consists of a near-field array of loudspeakers driven by a set of optimized signals in order to simulate random pressure fields with prescribed statistical properties over the surface of a panel. The grid of microphones defines the surface of simulation, which is a short distance away from the surface of a test panel.

During the simulation process, the loudspeakers are driven by a set of partially correlated optimal signals which are synthesized by a multichannel arbitrary wave form generator (AWG) system. The AWG system is illustrated by the block diagram shown in Fig. 3. The hardware includes a computer-based signal generator as well as a parallel-oriented array output processor. The signal generator enables the computing of the digital optimal signals $\mathbf{u}_{opt}$ in the frequency domain. It also provides the array output processor with the suitable time-domain signals $\tilde{\mathbf{u}}_{opt}$ by performing on each channel an IDFT of $\mathbf{u}_{opt}$. The length-N elements of the time domain vector $\tilde{\mathbf{u}}_{opt}$ are therefore given by

$$\tilde{u}_{opt,i}(m) = \frac{1}{N}\sum_{l=1}^{N} F_i(l) W_N^{-(l-1)(m-1)}, \quad m = 1, \ldots, N, \quad (10)$$

where the index i denotes the ith driving signal and $W_N = e^{-j2\pi/N}$. To assure real time domain signals, the length-$N/2$ elements of the frequency domain vector $\mathbf{u}_{opt}$ must be mirrored with proper symmetry such that, for each i,
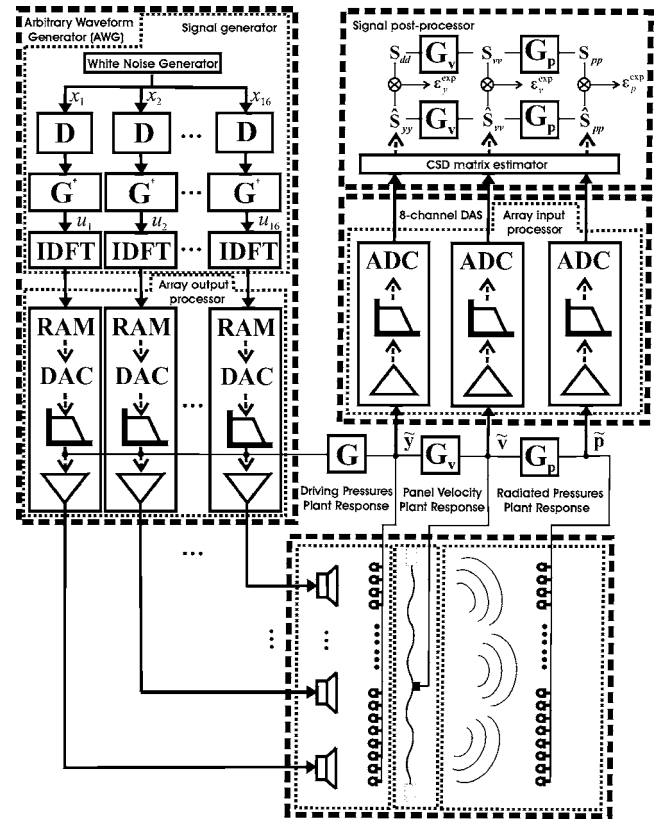
FIG. 3. Block diagram of the arbitrary wave form generator and the signal postprocessor that estimates the spatial errors associated to the excitation and to the panel vibroacoustic response.

$$\begin{cases} F_i(l) = u_{opt,i}^*(l), & l = 1, \ldots, N/2, \\ F_i(l) = u_{opt,i}(l - N/2), & l = N/2 + 1, \ldots, N. \end{cases} \quad (11)$$

The digital time-domain signals are then stored in a RAM memory organized as $32k$ words by 8 bits. The maximum number of samples per channel is limited to 32 767. When used in generation mode, a cyclic reading of the binary data can be performed simultaneously on all the memories. A synchronous counter output is also available to trigger the measurements on the real-time signals generation. The data on every channel are then passed through a digital-to-analog converter the output of which are filtered (5th order low-pass) to eliminate the clock residuals (see Fig. 3). The analog drive signals are the inputs to a set of 16 amplifiers delivering up to 50 W root-mean-square (rms) nominal power over 20 Hz–50 kHz bandwidth. The amplifier gain can be monitored on every channel, thus delivering 5, 10, 20, or 50 W rms power with an adjustable output voltage within each power range. A user interface enables to transfer the time samples to the memories, to start or stop the generation and to set up the sampling period of the drive signals (from 20 μs to 1 ms) as well as the cut-off frequency of the low-pass filters (from 2 kHz to 30 kHz).

2. Signal generation and acquisition

Prior to the experimental synthesis, a measurement of the plant matrix $\mathbf{G}$ is required and will be used to calibrate the system. For this purpose, the AWG system is used in

calibration mode and drives the amplifiers with sixteen un-correlated white noise signals with the following CSD matrix between the drive signals, $\mathbf{S}_{uu} = \mathbf{S}_{xx} = \mathbf{I}$. The CSD matrix $\mathbf{S}_{yu}$ between the microphone outputs $\mathbf{y}$ and the drive signals $\mathbf{u}$ allows the determination of the $\mathbf{G}$ matrix

$$\mathbf{S}_{yu} = E[\mathbf{yu}^{\mathrm{H}}] = E[\mathbf{Guu}^{\mathrm{H}}] = \mathbf{G}\mathbf{S}_{uu} = \mathbf{G}. \quad (12)$$

Each element of $\mathbf{G}$ corresponds to the frequency response function between one loudspeaker and one microphone. The frequency responses are estimated from time records of length $N=8192$ samples when collected over a duration $T=2.56$ s. Therefore, 4096 spectral lines were acquired for each frequency response at a sample rate $f_s=3.2$ kHz in order to cover frequencies up to $f_{\max}=N/2T=1.6$ kHz without temporal aliasing. For consistency, the same parameters have been kept for the acquisition of the time-domain microphones output signals, $\widetilde{\mathbf{y}}$, during the synthesis of the random pressure fields, thus setting to $T_A=T=2.56$ s the acquisition time of the synthesized signals at each microphone position.

The duration T_G of the drive signals generated during the synthesis process is set to a value greater than the acquisition time but at the expense of a reduced bandwidth that should cover the broadest frequency range over which the random pressure fields can be simulated. In a previous study it was found that the largest frequency range of simulation over which an acoustic diffuse field could *a priori* be reproduced stays below 1 kHz.[28] Therefore the generation bandwidth could be set up to 1.3 kHz, thus leading to a generation time $T_G=3.15$ s, still greater than T_A. In order to generate drive signals of length $N=8192$ samples over a duration T_G, a sampling period of 385 μs is required. We note that it is still lower than 1 ms, the maximum sampling period set by the memory cards. Also the drive signals are generated only from a single reading of the memories in order to avoid cyclic snapshots that may produce discontinuous time histories at the end sections of each block data.

During the synthesis, the optimal drive signals $\mathbf{u}_{\mathrm{opt}}$ are generated from a set of uncorrelated white noise reference signals $\mathbf{x}$ filtered by the optimal control matrix $\mathbf{W}_{\mathrm{opt}}$, as given by Eq. (5), such that $\mathbf{u}_{\mathrm{opt}}=\mathbf{W}_{\mathrm{opt}}\mathbf{x}$. In practice, the generated optimal signals are calculated off-line and implemented in the signal generator as $\mathbf{u}_{\mathrm{opt}}=\mathbf{G}^{\dagger}\mathbf{d}$. The assumed microphones outputs $\mathbf{d}$ are prefiltered by the inverse matrix $\mathbf{G}^{\dagger}$ in order to compensate for the effects of the plant response. During the simulation process, the target signals $\mathbf{d}$ should be recovered in a least-square sense at the microphones outputs $\mathbf{y}=\mathbf{G}\mathbf{u}_{\mathrm{opt}}$ if the plant response has been correctly equalized by the optimum control filter $\mathbf{W}_{\mathrm{opt}}=\mathbf{G}^{\dagger}\mathbf{D}$. The CSD matrix $\mathbf{S}_{yy}$ between the microphones outputs when subject to a synthesized random pressure field should then approximately match the required CSD matrix $\mathbf{S}_{dd}$. It reads

$$\mathbf{S}_{yy} = E[\mathbf{yy}^{\mathrm{H}}] = \mathbf{G}\mathbf{G}^{\dagger}\mathbf{S}_{dd}(\mathbf{G}\mathbf{G}^{\dagger})^{\mathrm{H}}. \quad (13)$$

The first equality of Eq. (13) is used to estimate the correlation structure between the experimentally synthesized pressures fields. It is obtained from the measured time-domain microphones outputs $\widetilde{\mathbf{y}}$ by means of averaging techniques, according to the Welch's method of smoothed periodo-

grams.[34] This method involves sectioning the length-N record data $\widetilde{\mathbf{y}}$ into M overlapping segments of length L, with D the offset between two consecutive segments so that $N=L+D(M-1)$, and averaging modified periodograms of these segments. An estimation of $\hat{\mathbf{S}}_{yy}$ from time records of length $N=8192$ samples is obtained with a reasonable trade-off between variance and spectral resolution when $\widetilde{\mathbf{y}}$ is subdivided into $M=255$ segments of length $L=2^6=64$ with 50% overlap. Therefore, the spectral density matrix $\mathbf{S}_{yy}$ is estimated as

$$\hat{\mathbf{S}}_{yy}(f_n) = \frac{L}{MN_w} \sum_{m=0}^{M-1} \mathbf{y}_{(m)}^*(f_n)\mathbf{y}_{(m)}^{\mathrm{T}}(f_n), \quad (14)$$

where $\mathbf{y}_{(m)}(f_n)=(1/L)\sum_{l=0}^{L-1}\widetilde{w}(l)\widetilde{\mathbf{y}}(l+mD)e^{-j2\pi lf_n}$ is the discrete Fourier transform (DFT) of the mth block of windowed data and $f_n=nf_s/L$, $n=0,\ldots,L/2$. The DFT of each block data is computed using a fast Fourier transform. $N_w=(1/L)\sum_{l=0}^{L-1}|\widetilde{w}(l)|^2$ is a normalization constant to compensate for the reduction in the average signal power induced by the truncation window $\widetilde{w}$.

III. EXPERIMENTAL FACILITY

The physical design of a loudspeakers array synthesis experiment is first described for the random pressure fields to be simulated over a test section. The modal characteristics of the test section, required to determine the sensitivity of the panel response to the simulated excitations,[29] are subsequently presented.

A. Design of the loudspeaker array synthesis experiment

An experimental device is built that implements the synthesis technique. It consists of a test panel clamped in a baffle and located inside a semianechoic chamber with dimensions 5 m $\times$ 5 m $\times$ 5 m. The desired pressure fields are reconstructed over a grid of electret microphones located in the proximity to the test panel of dimensions $l_x=0.74$ m, $l_y=0.92$ m. The pressures are generated at the microphone positions by a near-field array of loudspeakers mounted on a wooden enclosure of dimensions 1 m $\times$ 1 m $\times$ 0.2 m and centered at a distance h in front of the test section. The operating configuration of the experimental setup is shown in Figs. 2 and 4.

To obtain a characterization of the incident pressure field over a sufficiently dense grid of points, a set of 6-mm closely matched microphone capsules is chosen for their low cost, omnidirectional, and flat response performances in the operating frequency range, i.e., below 1 kHz. A detailed view of the sensor lay-out is presented in Fig. 5. One stainless steel bar is used for accurately positioning the microphones holders, with up to 15 evenly spaced locations along the bar to assess the effects of the number of microphones on the simulation accuracy. The bar acts like a baffle at high frequencies, causing a rise in gain only above 5 kHz. As it can be seen from Figs. 4 and 5, the bar is displaced along the panel length so that a maximum number of $N_x^{\mathrm{mic}} \times N_y^{\mathrm{mic}}=15 \times 18$ microphones positions are uniformly distributed over the
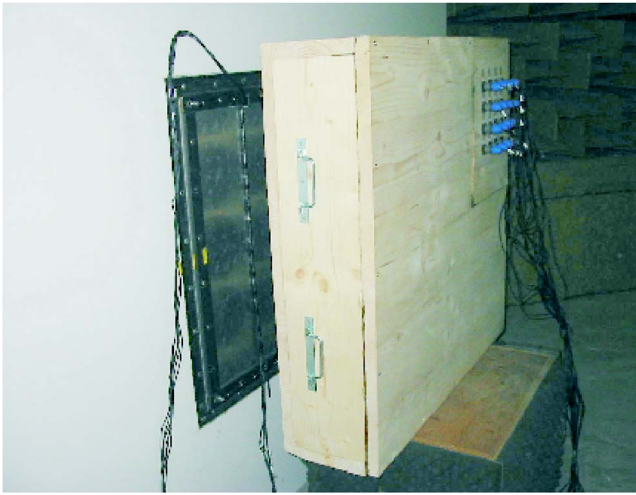
FIG. 4. The operating experimental setup.

FIG. 6. A front view of the loudspeakers cabinet.

panel surface, N_x^{mic} and N_y^{mic} being the number of microphones locations along, respectively, the width and the length of the panel. Due to the finite size of the microphones, the minimum distance between the sensors and the test panel is set to $\delta=0.01$ m.

The desired pressure field is generated by a set of $N_x^{loud} \times N_y^{loud} = 4 \times 4$ woofer loudspeakers of 0.21 m diameter uniformly distributed on the front face of the enclosure, as shown in Fig. 6. N_x^{loud} and N_y^{loud} are the number of loudspeakers along, respectively, the horizontal and vertical dimensions of the face of the enclosure. The number, the position

FIG. 5. A photograph of the microphone array layout.

and the diameter of the drivers were determined as a trade-off between the flatness of their frequency responses over the bandwidth of simulation, i.e., below 1 kHz, and their size limited by the surface of simulation, i.e., the panel dimensions. The loudspeakers cabinet is filled with fiberglass material to limit the internal acoustical coupling between the drivers and to damp the enclosure resonances, this technique being mostly effective above approximately 800 Hz. This uniform arrangement of loudspeakers might not be the most efficient configuration for reproducing a random pressure field with different correlation lengths along the two panel directions such as a TBL pressure field. However, it constitutes a convenient starting point.

The distance h between the array of loudspeakers and the test panel is optimized using a simplified analytical model in which the loudspeakers are modeled as acoustic monopoles radiating in front of a panel clamped in an infinitely rigid baffle. A mean-square error is defined as the ratio of the sum of the mean-square error signals at the microphones outputs to the sum of the mean square desired sensor signals. The minimum value for the normalized mean-square error is shown to be[27]

$$J_e = \frac{\text{Tr}[(\mathbf{I} - \mathbf{G}\mathbf{G}^\dagger)\mathbf{S}_{dd}]}{\text{Tr}[\mathbf{S}_{dd}]}. \tag{15}$$

As illustrated in Fig. 7, the ability of the transducers arrangement to reproduce the three types of random pressure fields is almost independent of the height of the loudspeakers provided it is greater than the loudspeakers separation distance. However, if the array of loudspeakers is too far from the test partition, the matrix of acoustic transfer functions $\mathbf{G}$ becomes ill-conditioned, thus leading to a poor reconstruction of the desired pressure field. Therefore, we have selected a suitable separation distance to the panel equal to the distance between two adjacent loudspeakers in the array, i.e., $h=0.22$ m. From this simplified model we already anticipate that the synthesis of a TBL pressure field will be much more demanding than for the other two types of excitations.

The truncation effects due to the limited aperture of both the loudspeakers and microphones arrays along each of the
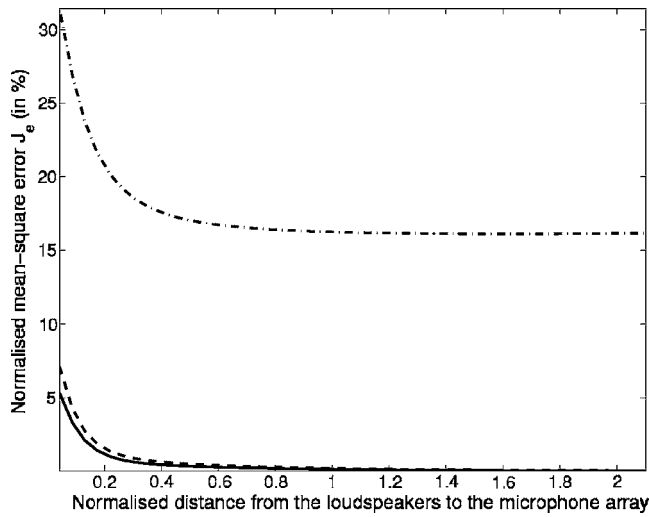
FIG. 7. Normalized mean-square error calculated at 120 Hz for the simulation of an acoustic diffuse field (solid), a grazing incident plane wave (dashed), and a TBL pressure field (dash-dotted) over an array of 15×18 microphones, as a function of the distance loudspeakers-microphones normalized by the separation distance between two adjacent loudspeakers.

FIG. 8. The condition number of the acoustic plant matrix measured between the 4×4 loudspeakers and a number of microphones (thin dashed line, 5×6 microphones; thick dashed line, 7×8 microphones; thin solid line, 13×16 microphones; thick solid line, 15×18 microphones).

panel dimensions should be considered. This limitation introduces a cutoff in the angular frequencies of the wave field radiated by the loudspeakers array at each microphone position. The largest angular bandwidth generated by the array is reached at the central microphone position and is given by $B = 2\sin[\tan^{-1}(a/2\tilde{h})]/\lambda$, where $a=1$ m is the aperture of the loudspeakers array and $\tilde{h} = h - \delta = 0.21$ m is the height of the loudspeakers array above the microphone plane. Shannon's sampling theorem defines the spatial resolution of the wave field at the microphone plane as $\delta_m = 1/B$. Assuming that the spatial resolution is limited by the separation distance between two microphones, i.e., $\delta_m = l_x/N_x^{mic} = l_y/N_y^{mic}$, then a synthesis can be carried out with this resolution and without spatial aliasing up to

$$f_{alias} = N_x^{mic} c/\{2 l_x \sin[\tan^{-1}(a/2\tilde{h})]\}, \tag{16}$$

i.e., up to about 3.2 kHz with 13×16 microphones and with a spatial resolution of 0.06 m.

The condition number of the plant matrix **G** plays an important role in order to explore how accurately the least-squares matrix of filters [Eq. (5)] is able to reconstruct the statistical properties of the desired pressure fields in a synthesis experiment. A set of 16×270 acoustic transfer functions has been measured between the loudspeakers and the microphones using an eight-channel LMS CADA-X acquisition system. It can be seen from Fig. 8 that the plant matrix condition number typically decreases as frequency increases. However, the values range well below 10^3, so there is *a priori* no need to use regularisation techniques in order to improve the reconstruction accuracy.[35] The variations with frequency of the plant matrix condition number result from the spectral behavior of all the measured transfer functions that takes into account the structural resonances of the baffled test panel, the semianechoic environment and the radiation properties of the loudspeakers cabinet. Testing on the cabinet alone has shown that the internal enclosure's depth
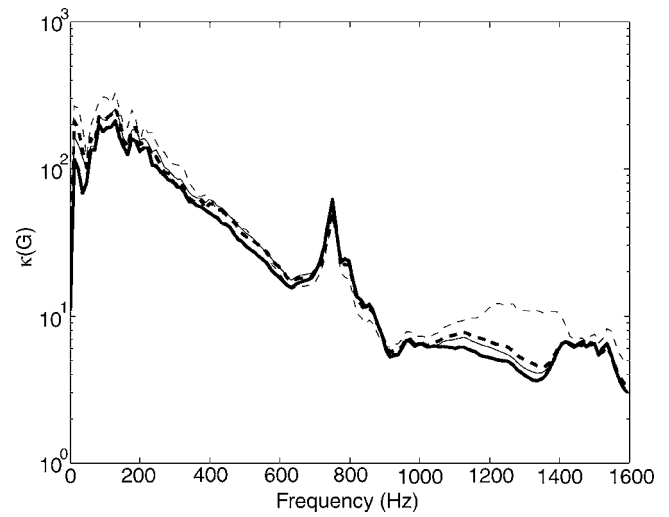
modes occur above 900 Hz, i.e., within a frequency range for which the dissipative material inside the cavity is effective. Also, the use of 5-cm thick plywood material for the cabinet walls increases the stiffness, mass, and damping of the enclosure, thus reducing the extent to which the cabinet vibrations contribute to the sound radiated. This design leads to a reasonably flat frequency response of the loudspeaker's array up to 1.6 kHz. However, when facing the baffled test section, local maxima are observed in the measured transfer functions. They occur at evenly spaced frequencies, $f_r = cr/2h$, $(r = 1, 2 \ldots)$, where c is the sound speed, and result from a regime of standing waves due to acoustical reflections between the baffled panel and the front face of the loudspeakers enclosure. The first acoustical resonance at 775 Hz dominates the spectrum of a large number of measured transfer functions. As seen from Fig. 8, it locally degrades the plant matrix condition number that still stays below 100 at this frequency. The second acoustical resonance does occur at 1550 Hz, but to a lesser extent.

The influence of the number of microphones on the plant matrix conditioning is also shown in Fig. 8. As expected, the matrix **G** is better conditioned when the density of microphones increases. A density of 7×8 microphones sets a lower limit above which the reproduction of the desired pressure fields should not be significantly improved. However, in a following paper, the results will be presented of pressure fields generated over an array of 13×16 microphones to provide enough spatial resolution on the reconstructed correlation structures.[29]

B. Vibration characteristics of the test section

To delineate error criteria induced by the simulated excitations on the test section, a model of the test panel structural behavior is required. As illustrated in Fig. 9, a modal analysis is performed on the clamped aluminium panel with thickness $h_p = 0.003$ m, with a density $\rho = 2650$ kg/m^3, a Young's modulus of 70.3 GPa, and a Poisson's ratio of 0.35.

T. Bravo and C. Maury: Random pressure fields synthesis
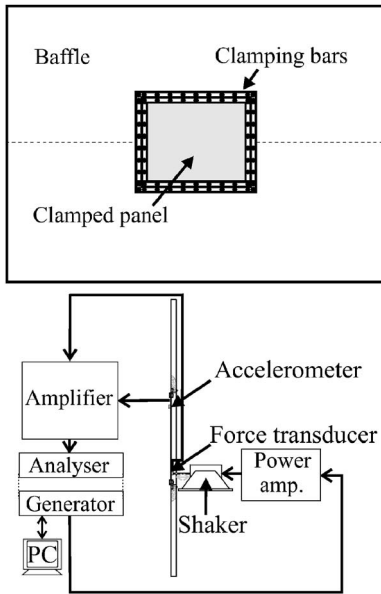
FIG. 9. Sketch of the experimental setup for modal testing.

FIG. 10. Experimental natural frequencies of the test panel vs the theoretical frequencies. For each frequency, the corresponding theoretical (left) and experimental (right) modal deflection shapes are compared.

The clamped boundary conditions are *a priori* well replicated in our experiment: the upper and lower sides of the panel edges are bolted together to eight stainless steel clamping bars, 0.05 m wide and 0.02 m thick, to prevent any transverse displacement and rotation of the edges. The clamping frame is mounted on a rigid fiberboard baffle.

An analytical solution for the panel response is derived using a classical modal approach. The structural velocity at a point **M** on the panel surface due to a punctual force located at **M′** denotes the transfer mobility function G_v. It can be expressed as a modal summation

$$G_v(\mathbf{M},\mathbf{M}';\omega) = \sum_{m,n=1}^{\infty} a_{mn}(\mathbf{M}';\omega)\psi_{mn}(\mathbf{M}), \qquad (17)$$

where a_{mn} is the complex amplitude of the mnth structural mode with normalized shape function ψ_{mn} and natural frequency ω_{mn}. Expressions are readily found for the normal modes and natural frequencies of a clamped panel.[36]

After substituting Eq. (17) into the variational equations of motion for the panel, one obtains the following expression for the modal amplitudes:

$$a_{mn}(\mathbf{M}';\omega) = -\frac{j\omega\psi_{mn}(\mathbf{M}')}{\rho h_p l_x l_y[(1+j\eta_{mn})\omega_{mn}^2 - \omega^2]}, \qquad (18)$$

where η_{mn} is the modal structural damping loss factor.

In order to validate the panel structural model, "single-point excitation" testing is performed (see Fig. 9).[37] Mobility frequency response functions are measured using a small-sized accelerometer, sequentially positioned at 17×21 points uniformly distributed, respectively, along the width and the length of the panel. A medium-sized electromagnetic driver with a 200 N force rating (sine peak) is attached to the rear part of the panel at a corner point of coordinates (x' =0.12 m, y'=0.17 m), which is not likely to coincide with the nodal lines of many structural modes. The shaker is linked to the panel via a drive rod, very stiff in the excitation direction, with impedance head to measure both force and acceleration at the excitation point.

From Fig. 10, it is clear that the first measured natural modes and frequencies are in good agreement with those calculated. It confirms that the experimental conditions are well modeled assuming a homogeneous panel with clamped edges. The structural damping factors of the panel are extracted by circle-fitting the Nyquist plots of the FRFs in the vicinity of each resonance.[38] After updating the modal damping in the model, a good correlation is obtained between the measured and calculated panel kinetic energies (see Fig. 11). The updated model is used in a following study[29] to assess the effect of the simulated excitations on the panel response through the introduction of spatial errors, $\varepsilon_y^{\exp}$, $\varepsilon_v^{\exp}$, and $\varepsilon_p^{\exp}$, respectively, for the excitation, the panel structural and acoustic responses, and calculated by the signal post-processor between the assumed and synthesized CSDs (see Fig. 3).
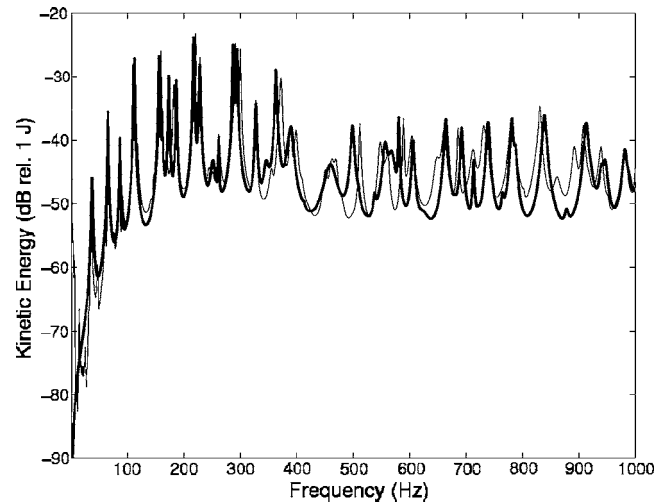
FIG. 11. Kinetic energy of the test panel when excited by a punctual force: theory (thick line) and experiment (thin line).

IV. CONCLUSIONS

A framework is presented for the multichannel simulation of spatially correlated random pressure fields with specified statistical properties. We aim to reproduce excitations with given spatial correlation properties such as an acoustic diffuse field, a grazing incident plane wave of random phase and amplitude, and the wall-pressure fluctuations beneath a high speed TBL. Other empirical or theoretical models can be considered for the correlation structure of the pressure field to be synthesized. The synthesis method requires the use of a near-field array of acoustic sources driven by a set of mutually correlated signals so that they reconstruct the spatial correlation structures of the desired pressure fields over a grid of microphones close to the test surface. The drive signals are the outputs of an optimal control filter which is designed off-line, prior to the synthesis experiment. During the synthesis, the drive signals are played out from an arbitrary wave form generator. The spatial correlation structures synthesized over the test section are estimated from measurements of the output microphones signals using averaging techniques. Unlike real-time reconstruction techniques, this approach does not require the causality of the optimal filters which operate on the assumed time histories. The optimal drive signals are obtained from the knowledge of both the target spectral density matrix factorized into a number of statistically uncorrelated components, and the matrix of physical responses between the acoustic sources and the microphones. These two factors *a priori* determine the physical limitation performances of the simulation process.

An experimental setup has been designed in a semi-anechoic chamber for the laboratory synthesis of the spatially correlated random pressure fields over a grid of 13 $\times$ 16 microphones when using a near-field array of 4×4 loudspeakers. It is found that the loudspeakers array, when driven by a set of optimal signals, should equalize for the acoustic transfer paths between the sources and the microphones provided that the height between loudspeakers and microphones is greater than the speakers separation distance. Moreover, the truncation effects due to the limited aperture of both the loudspeakers and microphones arrays sets an upper frequency limit of 3.2 kHz below which a synthesis can be carried out without spatial aliasing and with a resolution of 0.06 m.

At this design step, we anticipate that the synthesis of a TBL pressure field will be much more demanding than for the other two types of excitations. Such differences in the simulation accuracy with respect to the nature of the excitation is analyzed more in detail in a companion paper. In order to achieve the simulation of a TBL over a broader frequency range, further investigations should require a denser as well as a nonuniform arrangement of acoustic sources due to the different spanwise and streamwise correlation lengths involved. Also, refined models of TBL wall-pressure fluctuations should be considered, in particular those avoiding "corners" of equal reduction coherence as well as other models more suitable for low speed flows applications. *A priori*, the very short correlation lengths induced by low speed TBLs may require an unrealistic density of sources which would make the simulation very difficult to achieve. In general, quantitative criteria should be formulated on the number of acoustic sources required for an accurate simulation of each of the three random excitations.

[1] J. Mathys, "Low-frequency noise and acoustical standards," Appl. Acoust. **40**, 185–199 (1993).

[2] L. L. Beranek and I. L. Vér, *Noise and Vibration Control Engineering. Principles and Applications* (Wiley, New York, 1992).

[3] D. B. Pedersen, J. Roland, G. Raabe, and W. Maysenhölder, "Measurement of the low-frequency sound insulation of building components," Acust. Acta Acust. **86**, 495–505 (2000).

[4] M. Volander and A. C. C. Warnock, "Inter-laboratory comparisons of low frequency sound transmission: Conventional and intensity methods," *Proceedings of Internoise 93, Leuven*, Belgium, 1993, pp. 933–936.

[5] T. Kihlman, "Sound radiation into a rectangular room. Applications to airborne sound transmission in buildings," Acustica **18**, 11–20 (1967).

[6] K. A. Mulholland and R. H. Lyon, "Sound insulation at low frequencies," J. Acoust. Soc. Am. **54**, 867–878 (1972).

[7] L. Gagliardini, J. Roland, and J. L. Guyader, "The use of a functional basis to calculate acoustic transmission between rooms," J. Sound Vib. **143**, 457–478 (1991).

[8] P. S. Maluski and B. Gibbs, "Application of a finite-element model to low-frequency sound insulation in dwellings," J. Acoust. Soc. Am. **108**, 1741–1751 (2000).

[9] P. Santos and A. Tadeu, "A note on the acoustic insulation between two-dimensional acoustic spaces at low frequencies," J. Sound Vib. **261**, 185–191 (2003).

[10] T. Bravo and S. J. Elliott, "Variability of low frequency sound transmission measurements," J. Acoust. Soc. Am. **115**, 2986–2997 (2004).

[11] S. A. Rizzi and G. Bossaert, "Closed-loop control for sonic fatigue testing systems," Sound Vib. **35**, 19–23 (2001). It is also available online at the site ⟨http://hdl.handle.net/2002/15516⟩.

[12] A. Steinwolf, R. G. White, and H. F. Wolfe, "Simulation of jet-noise excitation in an acoustic progressive wave tube facility," J. Acoust. Soc. Am. **109**, 1043–1052 (2001).

[13] M. Shinozuka, "Simulation of multivariate and multidimensional random processes," J. Acoust. Soc. Am. **49**(2), 357–367 (1971).

[14] J. N. Yang, "Simulation of random envelope processes," J. Sound Vib. **21**, 73–85 (1972).

[15] W. Fricke, J. R. Bissel, and B. B. D'Ewart, "Acoustical fatigue problems associated with the design of a propeller duct," in *Acoustical Fatigue in Aerospace Structures*, edited by W. J. Trapp and D. M. Forney, Jr. (Syracuse University Press, New York, 1965), Chap. 33, pp. 685–705.

[16] W. V. Bhat, "Flight test measurements of exterior turbulent boundary layer pressure fluctuations on Boeing model 737 airplane," J. Sound Vib. **14**, 439–457 (1971).

[17] J. F. Wilby and F. L. Gloyna, "Vibration measurement of an airplane fuselage structure. I: Turbulent Boundary Layer excitation," J. Sound Vib. **23**, 443–466 (1972).

[18] W. H. Blake, "Structural response to turbulent wall flow and random sound," *Mechanics of Flow-Induced Sound and Vibration*, (Academic, New York, 1986) Vol. **17-II**, Chap. 9, pp. 595–654.

[19] M. J. Crocker, "The response of a supersonic transport fuselage to boundary layer and to reverberant noise," J. Sound Vib. **9**, 6–20 (1969).

[20] L. A. Schutzenhofer, "Beam stresses induced by a turbulent boundary layer and simulated with a reverberant acoustic field," *Proceedings of the AIAA Dynamics Specialist* Conference, Atlanta, United States, 1981 paper 81–0629, pp. 547–558.

[21] C. Maury, P. Gardonio, and S. J. Elliott, "A wavenumber approach to modelling the response of randomly excited panel, Part I: General theory,"

J. Sound Vib. **252**, 83–113 (2002).

[22]F. J. Fahy, "On simulating the transmission through structures of noise from turbulent boundary layer pressure fluctuations," J. Sound Vib. **3**, 57–81 (1966).

[23]G. Robert and J. Sabot, "Use of random forces to simulate the vibroacoustic response of a plate excited by a hydrodynamic turbulent boundary layer," *Proceedings of the ASME Winter Meeting: Symposium on Flow-Induced Vibrations, Vol. 5: Turbulence-Induced Noise and Vibration of Rigid and Compliant Surfaces*, 1984 pp. 53–61.

[24]A. Zerva and V. Zervas, "Spatial variation of seismic ground motions: An overview," Appl. Mech. Rev. **55**, 271–297 (2002).

[25]C. Maury, S. J. Elliott, and P. Gardonio, "A convergence study for the laboratory simulation of random pressure fields," on CD-ROM: *Proceedings of the Institute of Acoustics Spring Conference*, University of Salford, United Kingdom, 2002.

[26]C. Maury, S. J. Elliott, and P. Gardonio, "Turbulent boundary layer simulation with an array of loudspeakers," AIAA J. **42**, 706–713 (2004).

[27]S. J. Elliott, C. Maury, and P. Gardonio, "The synthesis of spatially correlated random pressure fields," J. Acoust. Soc. Am. **117**, 1186–1201 (2005).

[28]T. Bravo and C. Maury, "Experimental reproduction of random pressure fields in laboratory conditions," J. Acoust. Soc. Am. **117**(2), 2547 (2005).

[29]C. Maury and T. Bravo, "The experimental synthesis of random pressure fields: Practical feasibility," J. Acoust. Soc. Am. **120**(5) (in press).

[30]S. J. Elliott, *Signal Processing for Active Control* (Academic, London, 2000).

[31]F. Jacobsen, "The diffuse sound field," The Acoustics Laboratory, Technical University of Denmark, Report No. 27 (1979).

[32]G. M. Corcos, "Resolution of pressure in turbulence," J. Acoust. Soc. Am. **35**, 192–199 (1963).

[33]D. M. Chase, "The character of the turbulent wall pressure spectrum at subconvective wavenumbers and a suggested comprehensive model," J. Sound Vib. **112**, 125–147 (1987).

[34]P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans. Audio Electroacoust. **AU-15**, 70–73 (1967).

[35]S. H. Yoon and P. A. Nelson, "Estimation of acoustic source strength by inverse methods: Part II, experimental investigation of methods for choosing regularization parameters," J. Sound Vib. **233**, 669–705 (2000).

[36]J. P. Arenas, "On the vibration analysis of rectangular clamped plates using the virtual work principle," J. Sound Vib. **266**, 912–918 (2003).

[37]D. J. Ewins, *Modal Testing: Theory, Practice and Applications* (Research Studies, Baldock, Hertfordshire, England, 2000).

[38]J. Piranda, "Analyse modale et recalage de modèles," HDR Project (Habilitation à Diriger des Recherches), Université de Franche-Comté, Besançon, France, 1994.

The experimental synthesis of random pressure fields: Practical feasibility

Cédric Maury and Teresa Bravo[a)]
Laboratoire Roberval FRE-CNRS 2833, Université de Technologie de Compiègne, Secteur Acoustique,
Boîte Postale 20529, F-60205 Compiègne Cedex, France

A previous paper discussed the methodology for the synthesis of partially correlated random pressure fields using a near-field array of loudspeakers. The acoustic sources are optimally driven so that various random excitations are reproduced over a test surface, namely an acoustic diffuse field, a grazing incident plane wave, and turbulent boundary layer fluctuating loads. This paper shows the physical limitation performances and the practical feasibility of synthesizing these random pressure fields in a series of loudspeakers array simulation experiments. Spatial error criteria are proposed on the number of acoustic sources per unit correlation length. They are more representative than mean-square error criteria to quantify the accuracy with which the assumed correlation structures are experimentally reconstructed. Furthermore, structural and acoustic models are formulated to investigate how sensitive is the panel vibroacoustic response to inaccuracies in the synthesized excitations. It is discussed how the direct reproduction of the panel vibroacoustic response with a limited number of loudspeakers should be feasible within the frequency bandwidth for which the modes at resonance well couple with the excitation. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354005]

I. INTRODUCTION

The problem of simulating multivariate and multidimensional random processes with prescribed spectral density characteristics has been extensively addressed in the past decades both theoretically and experimentally for the synthesis of real-world acoustic or vibration environments. A review about the methodologies used to synthesize random pressure fields has been described in a previous work.[1] These methods have been assessed within different areas for the experimental synthesis of partially correlated signals. For example, a parametric approach based on the simulation of pseudorandom time histories has been successfully applied in the automotive industry for the laboratory reproduction of the response of a four-track vehicle to the Gaussian excitation induced by road surface roughness.[2,3] It has also been used for simulating the spatial variation of seismic ground motions recorded over extended areas.[4] More recently, this method has been applied for tailoring in-service environmental vibrations of ground vehicles with non-Gaussian characteristics using shaker testing facilities.[5]

In the frequency domain, an equivalent approach consists in factorizing the spectral density matrix of the target data in order to design shaping filters the output of which are the simulated spectra.[6] It is well adapted for the experimental simulation of multidimensional random fields such as boundary-layer turbulence or acoustic diffuse fields. In particular, a point array forcing experiment (PAFE) has been designed in order to achieve cost-efficient "ground testing" to reproduce turbulent boundary layer (TBL) noise transmis-

sion through aircraft sidewalls.[7] An array of suitably driven shakers was used to reproduce the statistics of the TBL wall-pressure fluctuations over an aluminium panel. The PAFE technique was found to be well suited to simulate resonant forcing of the panel like "rain-on-the-roof" excitation with respect to nonresonant excitations such as those due to a TBL or an acoustic diffuse field. One way of reproducing nonresonant excitations would be to excite the panel by a suitably correlated sound field as it could be generated using a near-field array of loudspeakers. Such an approach builds upon previous theoretical works concerning the synthesis of TBL wall-pressure fluctuations using a number of suitably driven acoustic monopoles.[8,9] Subsequently, an experimental setup has been designed for the laboratory reproduction of spatially correlated random pressure fields with a near-field array of 4×4 loudspeakers.[1,10] Preliminary investigations have shown that this arrangement should *a priori* be able to reproduce the statistics of an acoustic diffuse field and a grazing incident plane wave over a broad frequency range whereas the simulation of TBL loads appears to be restricted to the very low frequency domain.

The present work addresses practical issues for the laboratory synthesis of three random pressure fields, namely an acoustic diffuse field, a grazing acoustic plane wave, and a TBL excitation. The accuracy that can be experimentally achieved for the reconstruction of each kind of excitations is quantified with respect to both energy-based and spatial errors, as explained in Sec. II. In Sec. III, the physical limitations of the simulation performance are assessed and criteria are proposed on the number of sources required for an accurate reconstruction of the excitations. Section IV discusses the experimental results on the off-line synthesis of each of

---

[a)]Present address: iTEAM, Universidad Politècnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain. Electronic mail: tebrama@iteam.upv.es

the three random pressure fields. In Sec. V, results are given on the sensitivity of the panel vibroacoustic response to inaccuracies in the simulated excitations.

II. QUANTIFIERS OF THE SIMULATION ACCURACY

In order to implement the synthesis technique, a possible experimental facility has been designed and constructed for the evaluation of test cases.[1] Briefly, the excitations are reconstructed over a grid of $N_x^{mic} \times N_y^{mic} = 13 \times 16$ microphones evenly distributed, respectively, along the width and the length of a baffled clamped panel located inside a semi-anechoic chamber. The random pressure fields are generated using a near-field array of 4×4 loudspeakers, suitably driven and located at a distance $\tilde{h} = 0.21$ m apart from the grid of microphones which defines the surface of simulation. The height, $\tilde{h}$, between the loudspeakers and the microphones grid was chosen, typically about the separation distance between two adjacent loudspeakers, in order to lower the condition number of the plant response matrix $\mathbf{G}$ between the loudspeakers and the microphones. This number is an indicator of the ill-posedness of the problem associated with the pseudoinversion of $\mathbf{G}$, required for the off-line calculation of the loudspeakers driving signals. Also, given the limited aperture of the loudspeakers array and the spatial resolution set by the microphones separation distance, it has been shown that synthesis can be performed up to about 3.2 kHz without spatial aliasing.[1]

A. The mean-square error

From the knowledge of both the statistical properties for the desired pressure fields, characterized by the spectral density matrix $\mathbf{S}_{dd}$, and the measured transfer functions constitutive of the plant matrix $\mathbf{G}$, a normalized mean-square error is defined to quantify the accuracy of the simulated excitation. It is obtained as the sum of the mean-square error signals at the microphones outputs normalized by the sum of the mean square target sensor signals. Its minimum value is achieved when the loudspeakers are driven by a set of optimal signals. It is given by[9]

$$J_e = \frac{\text{Tr}[(\mathbf{I} - \mathbf{G}\mathbf{G}^\dagger)\mathbf{S}_{dd}]}{\text{Tr}[\mathbf{S}_{dd}]}, \qquad (1)$$

where Tr is the trace operator and $\mathbf{G}^\dagger$ is the pseudoinverse of $\mathbf{G}$.

Another aspect of the simulation is to assess the accuracy with which a reduced set of loudspeakers can approximate the panel acoustic response. A normalized mean-square error, J_p, is defined to quantify the approximation induced on the acoustic response of the panel when subject to the simulated random pressure fields

$$J_p = \frac{\text{Tr}[\mathbf{G}_p\mathbf{S}_{ee}\mathbf{G}_p^H]}{Tr[\mathbf{G}_p\mathbf{S}_{dd}\mathbf{G}_p^H]}, \qquad (2)$$

where H denotes the Hermitian, complex conjugate transpose; $\mathbf{S}_{ee} = [\mathbf{G}\mathbf{G}^\dagger - \mathbf{I}]\mathbf{S}_{dd}[(\mathbf{G}\mathbf{G}^\dagger)^H - \mathbf{I}]$ denotes the spectral density matrix for the error signals at the microphones outputs when the loudspeakers are optimally driven; $\mathbf{G}_p$ is

the transfer impedance matrix the elements of which are calculated from the Green's representation of the acoustic pressure radiated and associated to a given panel velocity distribution [see Eq. (A5)]. The velocity is calculated by the updated model presented in Sec. III B. of Ref. 1.

B. The residual error

Equation (1) is a discrete version of the normalized power spectral density averaged over the microphones locations and associated to the error signals at the microphones outputs. Hence, the mean-square error quantifies the accuracy of the approximation in terms of the incident power associated to the error signals and it is not fully representative of the accuracy with which the correlation structures are reproduced. A quantifier that emphasises the spatial characteristics of the random field statistics is defined as

$$\varepsilon_y^{exp} = \frac{\|\mathbf{C}_{dd} - \hat{\mathbf{C}}_{yy}\|}{\|\mathbf{C}_{dd}\|}, \qquad (3)$$

where $\|\cdot\|$ is the Frobenius norm. It is associated to the residual matrix $\mathbf{C}_{dd} - \hat{\mathbf{C}}_{yy}$ between a specified correlation matrix $\mathbf{C}_{dd}$ and the one estimated from experimental synthesis, $\hat{\mathbf{C}}_{yy}$. A residual error, ε_y, can also be defined with respect to the correlation matrix, $\mathbf{C}_{yy}$, predicted from the measured transfer functions. The normalized spatial correlation matrix, $\hat{\mathbf{C}}_{yy}$, is calculated with respect to a reference signal, $\tilde{y}_{j_0}$, associated to a specific microphone in the array, with entries

$$\hat{C}_{yy}(i,j_0) = \frac{\hat{S}_{yy}(i,j_0)}{\sqrt{\hat{S}_{yy}(i,i)\hat{S}_{yy}(j_0,j_0)}}, \qquad (4)$$

so that $|\hat{C}_{yy}(i,j_0)| \leq 1$ for all $i = 1, \ldots, N_x^{mic} \times N_y^{mic}$. $\hat{S}_{yy}(i,j)$ denotes the cross-spectral density (CSD) function synthesized between the ith and the jth microphone output signals, respectively $\tilde{y}_i$ and $\tilde{y}_j$ for $i,j = 1, \ldots, N_x^{mic} \times N_y^{mic}$. In the following, the reference signal is by default associated to a microphone located at the center of the panel, but it could also be associated to any microphone in the array. Indeed, the CSD functions only depend on the distance between any pair of points for all the microphone locations: they are identically and simultaneously produced with respect to each of the reference microphone which could be at any location in the array. A similar remark also applies to $\mathbf{C}_{yy}$.

We note that the residual error, ε_y (or ε_y^{exp}), sets an upper bound on the relative error due to the differences between the correlation area of the assumed and synthesized random pressure fields, respectively A_d and A_y, such that

$$\varepsilon_y^2 = \frac{A_{d-y}}{A_d} \geq \frac{|A_d - A_y|}{A_d}, \qquad (5)$$

due to the triangle inequality satisfied by the Frobenius norm. A_y is defined in terms of the normalized correlation function C_{yy} as
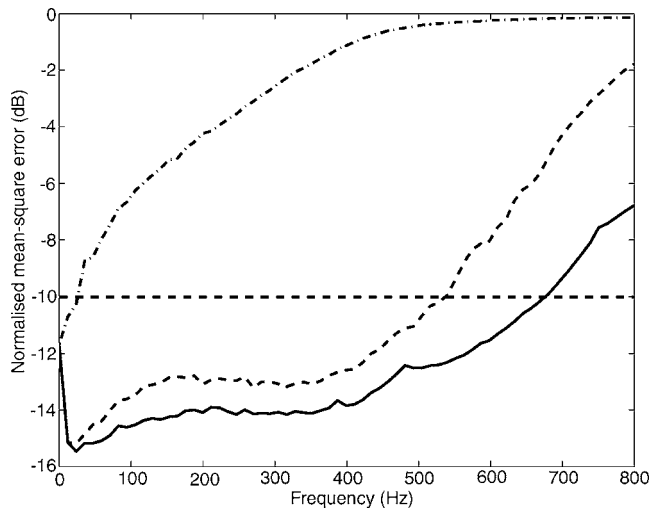
FIG. 1. Normalized mean-square error for the simulation of an acoustic diffuse field (solid), a grazing plane wave (dashed), and a TBL pressure field (dash-dotted).

$$A_y = \frac{\displaystyle\int\!\!\int_{\infty} |C_{yy}(r_x, r_y; \omega)|^2 dr_x dr_y}{|C_{yy}(0,0;\omega)|^2}, \qquad (6)$$

which can be approximated as $l_x l_y \|\mathbf{C}_{yy}\|^2 / (N_x^{\text{mic}} \times N_y^{\text{mic}})$ after discretization of Eq. (6) over a grid of microphones covering the surface of simulation.

The accuracy with which the spatial correlation structures of the panel response are approximated is quantified by the residual errors, $\varepsilon_\nu^{\exp}$ and $\varepsilon_p^{\exp}$, respectively, associated to the structural and acoustic responses of the panel [see Fig. 3 in Ref. 1]. They are defined as

$$\varepsilon_\alpha^{\exp} = \frac{\|\mathbf{C}_{\alpha\alpha} - \hat{\mathbf{C}}_{\alpha\alpha}\|}{\|\mathbf{C}_{\alpha\alpha}\|}, \qquad \alpha \in \{\nu, p\}, \qquad (7)$$

in terms of the assumed and synthesized spatial correlation matrices, respectively $\mathbf{C}_{\alpha\alpha}$ and $\hat{\mathbf{C}}_{\alpha\alpha}$.

III. PHYSICAL LIMITATIONS OF RANDOM PRESSURE FIELDS SYNTHESIS

An estimate is predicted of the frequency range over which the statistical properties of each kind of target pressure field are reproduced with accuracy by the array of loudspeakers (see Fig. 1). Previous studies[8,9] have shown that a −10 dB mean-square error J_e can be considered as a threshold below which an accurate reproduction of the desired pressure field is achieved. For completeness, comparisons are also made between the assumed and predicted spatial correlation structures, $\mathbf{C}_{dd}$ and $\mathbf{C}_{yy}$, for each of the three random pressure fields to be synthesized (top and middle rows of Figs. 2–4).

It can be seen from Fig. 1 that accurate simulations of an acoustic diffuse field and a grazing incident plane wave can be achieved, respectively, up to about 680 and 550 Hz, when using a near-field array of 4×4 loudspeakers. Over these frequency ranges, the optimal matrix of control filters is able to equalize the acoustic transfer functions measured between the loudspeakers and the microphones. However, another factor limits the performance of the random field synthesis: the number of sources required per correlation length to achieve an accurate reconstruction of each type of excitations over the panel. This factor is shown to be critical for the TBL pressure field statistics.

Empirical energy-based criteria are obtained for an accurate reproduction in terms of the number of sources per
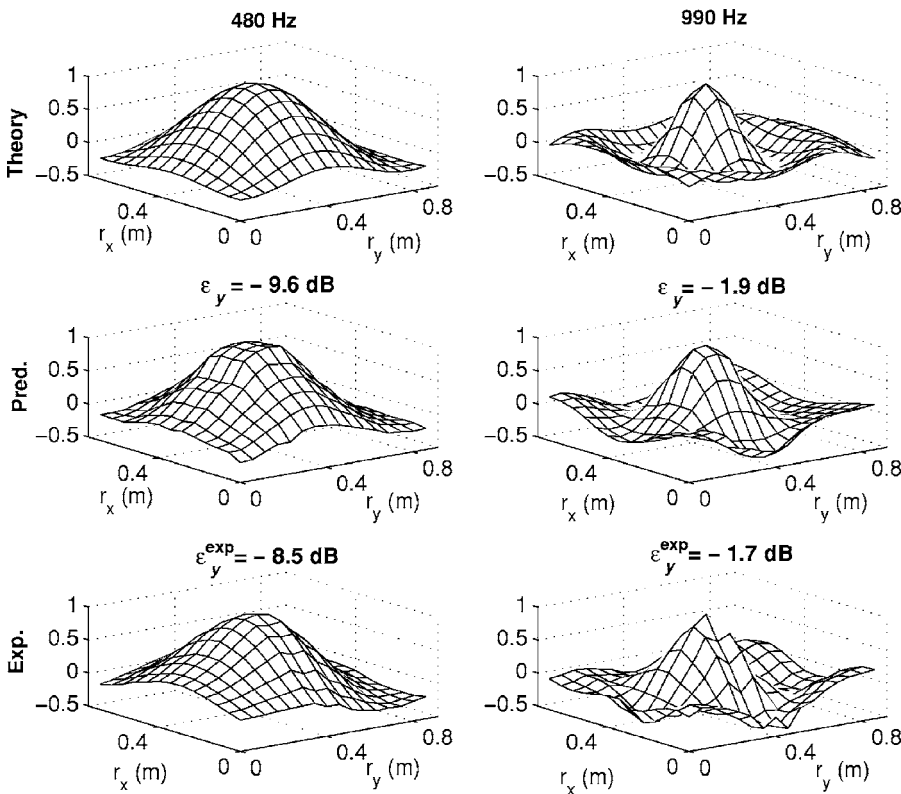
FIG. 2. Spatial correlation structures at 480 (left) and 990 Hz (right) when perfect reproduction of an acoustic diffuse field is assumed (top), that predicted from the measured plant responses (middle), and that experimentally synthesized with 4×4 loudspeakers (bottom).

C. Maury and T. Bravo: Experimental synthesis of random pressure fields
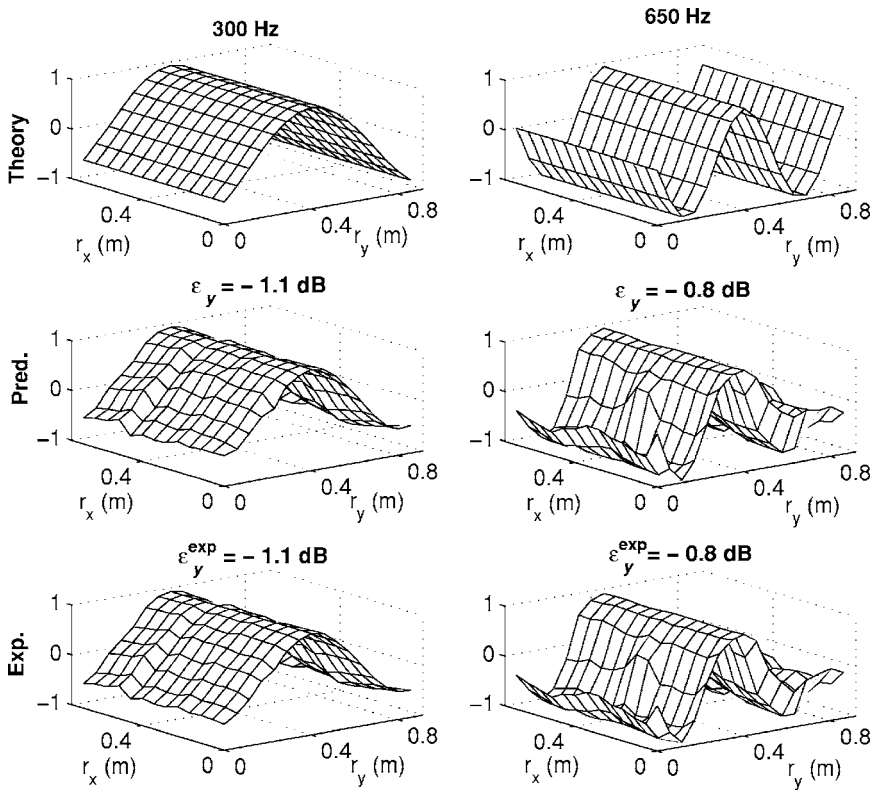
FIG. 3. Spatial correlation structures at 300 (left) and 650 Hz (right) when perfect reproduction of an acoustic plane wave is assumed (top), that predicted from the measured plant responses (middle), and that experimentally synthesized with 4×4 loudspeakers (bottom).

unit correlation length along each of the panel directions, i.e., in terms of $N_x^{\text{loud}} L_x^{\text{cor}}/l_x \times N_y^{\text{loud}} L_y^{\text{cor}}/l_y$, where $N_x^{\text{loud}} \times N_y^{\text{loud}}$, $L_x^{\text{cor}} \times L_y^{\text{cor}}$, and $l_x \times l_y$ are, respectively, the number of sources, the correlation lengths, and the dimensions along each of the panel directions. They are lower bounded by theoretical criteria based on the proper orthogonal decomposition (POD) of the corresponding random processes which leads to the minimum number of uncorrelated components required for the mean-square approximation of each type of excitations. Note that, in order to make the theoretical problem tractable, it is assumed that the total required number of components is equal to the product of the required number of components along each of the panel directions.

A. Acoustic diffuse field

It is deduced from Fig. 1 that 2.7×2.2 acoustic sources per unit acoustic wavelength are required along the panel x and y directions for an accurate reproduction of a diffuse pressure field, thus setting the upper frequency performance of our transducers configuration at 680 Hz. The assumed CSD matrix for an acoustic diffuse field, as given by Eq. (6) in Ref. 1, is associated to a nonrational bandlimited random process. The mean-square values of the POD coefficients, which are also the eigenvalues λ_{mn} of $\mathbf{S}_{dd}$, are given by[11]

$$\lambda_{mn}(k) = \mathrm{R}_{0m}\left(\frac{kl_x}{2},1\right)^2 \mathrm{R}_{0n}\left(\frac{kl_y}{2},1\right)^2, \tag{8}$$

where R_{0m} is known as a radial prolate spheroidal wave function of the first kind and k is the acoustic wave number. The eigenvalues λ_{mn} rapidly decay and approach zero for values of m and n higher than kl_x/π and kl_y/π, respectively. Therefore, for any given k, only a finite number of eigen-

functions or POD components optimally represents the CSD matrix for an acoustic diffuse field. An analytical criterion can thus be determined on the number of uncorrelated components that best reproduce a given acoustic diffuse field over the surface of a test panel. This number scales with the squared frequency and is given by

$$N_{\text{Diff.}} = \frac{(kl_x)(kl_y)}{\pi^2}. \tag{9}$$

The theoretical criterion (9) corresponds to two sources per unit acoustic wavelength. Alternatively, for a given number of sources, $N_{\text{Diff.}}$, this criterion provides an upper frequency limit $f_{\text{Diff.}}$ for the approximation of an acoustic diffuse field

$$f_{\text{Diff.}} = \frac{c}{2}\sqrt{\frac{N_{\text{Diff.}}}{l_x l_y}}, \tag{10}$$

where c is the sound speed. Assuming 16 sources, the criterion (10) predicts a theoretical upper bound of 825 Hz for the approximation of an acoustic diffuse field, which is 17% above the maximum frequency of 680 Hz, obtained from the -10 dB mean-square error criterion.

In Fig. 2, we note an excellent reproduction of the diffuse field spatial correlation functions at 480 Hz with a residual error of -9.6 dB. We observe that the assumed and predicted correlation structures stay reasonably similar at 990 Hz, i.e., well above the frequency range of simulation initially determined from the -10 dB threshold, thus suggesting that the energy-based quantifier J_e might lead to rather stringent criteria with respect to the residual error, ε_y.
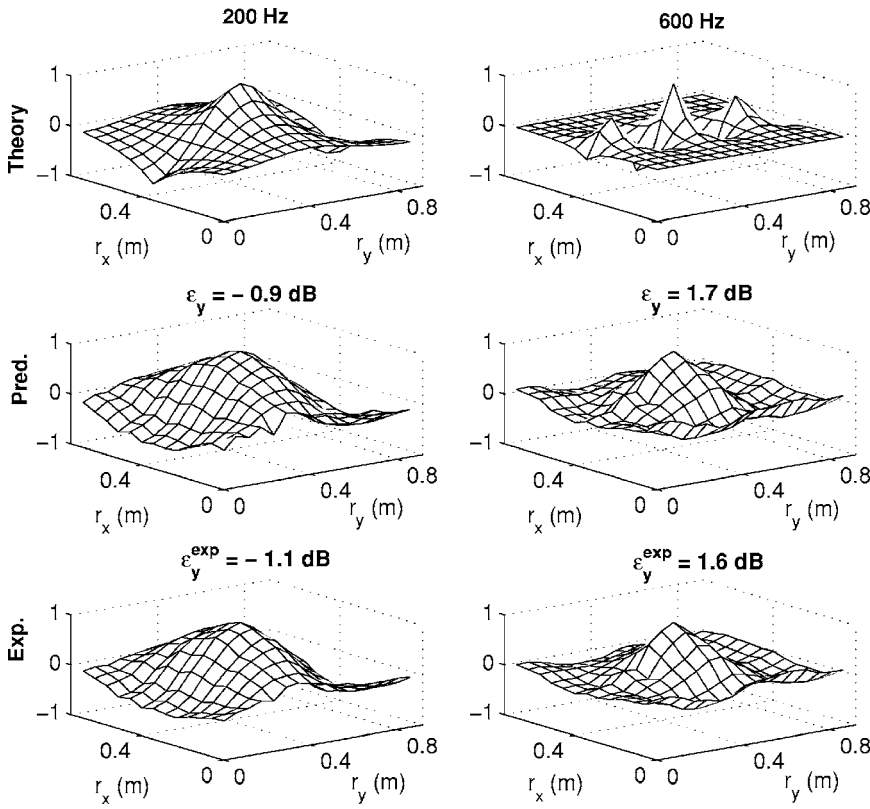
FIG. 4. Spatial correlation structures at 200 (left) and 600 Hz (right) when perfect reproduction of TBL is assumed (top), that predicted from the measured plant responses (middle), and that experimentally synthesized with 4×4 loudspeakers (bottom).

B. Grazing incident plane wave

The POD decomposition of a periodic homogeneous random field, such as a plane wave of random phase and amplitude [see Eq. (7) in Ref. 1], only requires one independent component, which is a Fourier mode with a wavelength scaling on the spatial period of the correlation function. This satisfies the optimality property of the POD basis functions. For two-dimensional processes, the Fourier mode has the form $e^{2\pi j(x/\Lambda_x + y/\Lambda_y)}$ where $\Lambda_x = \lambda / \sin\theta_K \sin\varphi_L$ and $\Lambda_y = \lambda / \sin\theta_K \cos\varphi_L$ are the traces of the acoustic wavelength λ along each of the panel directions, θ_K and φ_L are the polar and azimuthal incidence angles.

Figure 1 shows that experimental reproduction of a grazing incident plane wave can be attempted up to about 550 Hz which corresponds to 2.7 sources per unit acoustic wavelength along the y direction. More generally, given a frequency range for the simulations, empirical criteria on the number of sources per unit acoustic wavelength required to reproduce a general incident plane wave are found to be

$$N_x^{\mathrm{PW}} \approx 3.4 \left(\frac{kl_x}{2\pi} \right) \sin\theta_K \sin\varphi_L, \tag{11a}$$

$$N_y^{\mathrm{PW}} \approx 2.7 \left(\frac{kl_y}{2\pi} \right) \sin\theta_K \cos\varphi_L. \tag{11b}$$

We note that the theoretical criterion only predicts 0.7 sources per unit acoustic wavelength along the y direction and therefore greatly underestimates the empirical criterion (11b) which mainly reflects the ability of the transducers array to equalize for the acoustical plant response. Further simulations have been achieved when θ_K and φ_L are varied, confirming the validity of the above empirical criteria. In

particular, assuming $\theta_K = 45°$ and $\varphi_L = 0°$, the frequency range of simulation with the loudspeakers array extends up to 700 Hz which is similar to the frequency bandwidth over which an acoustic diffuse field can be well reproduced. The approximate correlation functions shown in Fig. 3 at 300 and 650 Hz correspond, respectively, to −1.1 and −0.8 dB residual error. The comparisons confirm the ability of the loudspeakers configuration to achieve a reasonably good spatial approximation of the grazing incident plane wave over a frequency range wider than the one initially set by the mean-square error criterion, i.e., up to about 550 Hz.

C. TBL pressure field

The CSD matrix for the wall-pressure fluctuations induced by a high speed TBL, as given by the Corcos model [see Eq. (8) in Ref. 1], is associated to a rational nonband-limited random process. The mean-square values, $\lambda_{n,x}$ and $\lambda_{n,y}$, of the POD decomposition of the TBL random processes, along, respectively, the spanwise and the streamwise directions, read

$$\lambda_{n,x} = \frac{(k_c l_x)^2}{(k_c l_x)^2 + [\alpha_x(n-1)\pi]^2}, \quad n = 1, 2, \ldots, \tag{12a}$$

$$\lambda_{n,y} = \frac{(k_c l_y)^2 (1 + \alpha_y^2)}{(k_c l_y)^2 + \alpha_y^2 [(n-1)\pi - k_c l_y]^2}, \quad n = 1, 2, \ldots, \tag{12b}$$

where $k_c = \omega / U_c$ is the convective wave number, U_c is the convection velocity and α_x, α_y are empirical parameters for the Corcos model. The components (12a) and (12b) are calculated as the solutions of a homogeneous differential equa-

tion which, for a rational spectrum, is equivalent to the POD integral eigenvalue problem whose kernel is the random process correlation function.[12] The eigenvalues are monotonically decreasing functions of n and lower bounds can be derived on the number of uncorrelated components, N_x^{TBL} and N_y^{TBL}, that capture the dominant dynamics of the random process. Given a threshold $T(\text{dB}) = 10 \log_{10}(J_e)$ on the average energy captured by the first POD components and after integrating (12a) and (12b) over the component index n, it is found that

$$N_x^{\text{TBL}} = \frac{k_c l_x}{\alpha_x \pi} \tan\left[(1 - 10^{T/10}) \frac{\pi}{2} \right], \tag{13a}$$

$$N_y^{\text{TBL}} = \frac{k_c l_y}{\alpha_y \pi} \left\{ \alpha_y + \tan\left[(1 - 10^{T/10}) \frac{\pi}{2} - 10^{T/10} \tan^{-1}(\alpha_y) \right] \right\}. \tag{13b}$$

Equation (13a) is the limiting criterion on the number of sources required in the spanwise direction since the spanwise correlation length is about eight times shorter than the streamwise correlation length. Given a -10 dB mean-square error, Eq. (13a) predicts 2.1 acoustic sources required per unit spanwise correlation length. With an array of 4×4 sources, it corresponds to upper theoretical frequencies of 92 and 297 Hz for the reproduction of a TBL pressure field induced by a high speed flow, respectively, along the x and y directions.

From Fig. 1, an accurate experimental reproduction of the TBL for high speed flows can only be achieved below about 50 Hz which corresponds to an empirical criterion of 3.7 sources per unit spanwise correlation length. As expected, the -10 dB criterion is lower bounded by the theoretical criterion. Figure 4 shows that the assumed and simulated correlation structures are still quite similar at 200 Hz, i.e., well above the upper frequency range of simulation predicted from the -10 dB criterion in the spanwise direction. From the above criteria, the simulation of TBL excitations induced by low speed flows is found to be very difficult to achieve even at low frequencies, due to the high number of correlation lengths to be synthesized over the panel surface as frequency increases. It would require a much denser array using necessarily smaller acoustic sources with limited performances at low frequencies.

IV. RESULTS ON THE EXPERIMENTAL SYNTHESIS OF THE EXCITATIONS

A. Influence of the loudspeakers configurations

Prior to conducting the synthesis experiments, the practical influence is analyzed of both the number and the location of the loudspeakers on the accuracy with which an acoustic diffuse field is reproduced. The conclusions drawn for the diffuse field case can also be applied to the other types of excitations. The results presented in Fig. 5 correspond to the mean-square error reduction, calculated from the measured transfer functions between a number of loudspeakers and an array of 13×16 microphones. Sets of either 3×3 or 3×4 source array configurations, as subsets of the
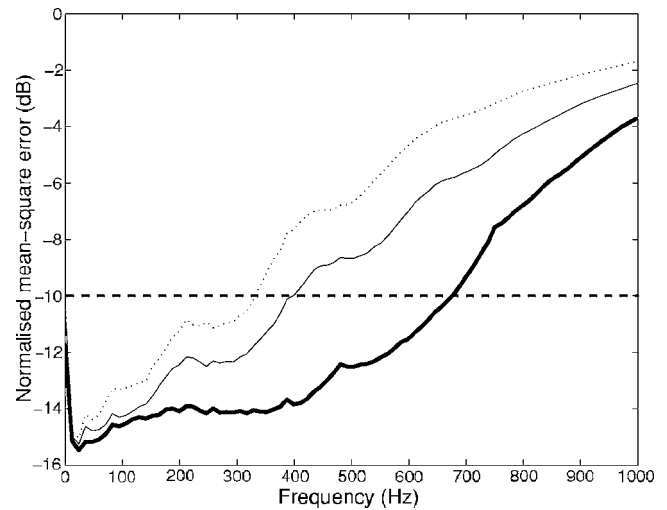
FIG. 5. Normalized mean-square error calculated from the measured plant matrix when 3×3 (dotted), 3×4 (thin solid), and 4×4 (thick solid) loudspeakers were used to reproduce an acoustic diffuse field.

4×4 array initially designed for the experiment, have been considered. As expected, the frequency range for an accurate simulation of a random pressure field increases with the number of loudspeakers used within the simulation area. Apart from the number of loudspeakers, the displacement of the sources array from a symmetric position relative to the microphones grid, i.e., the eccentricity, also contributes to degrade the reconstruction accuracy via the plant matrix condition number, but to a lesser extent. This is particularly so for a large number of microphones.

An acoustic diffuse field is experimentally synthesized using the arbitrary wave form generator (AWG) described in Sec. II B of Ref. 1 for each of the source array configurations described above. Figure 6 shows the variations of the corresponding residual error, as given by Eq. (3), when estimated from the synthesized random pressures measured at the microphones positions.[1] It can be seen that the correlation structures are generally better reconstructed as the number of loudspeakers increases and as the geometry of the source
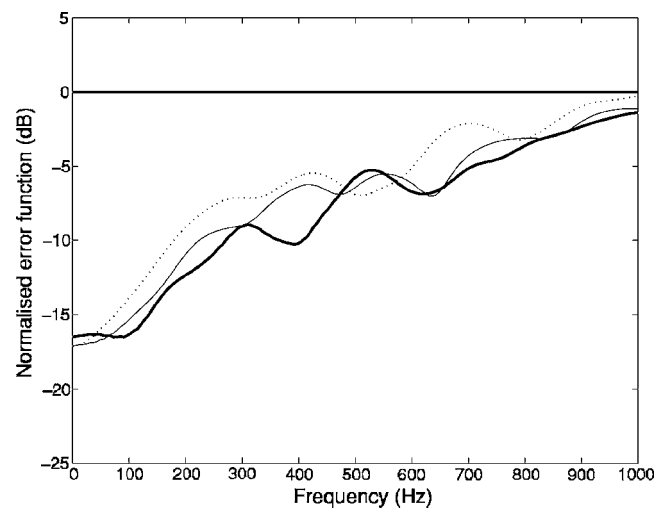
FIG. 6. Normalized residual errors estimated when 3×3 (dotted), 3×4 (thin solid), and 4×4 (thick solid) loudspeakers were used for the experimental synthesis of an acoustic diffuse field.
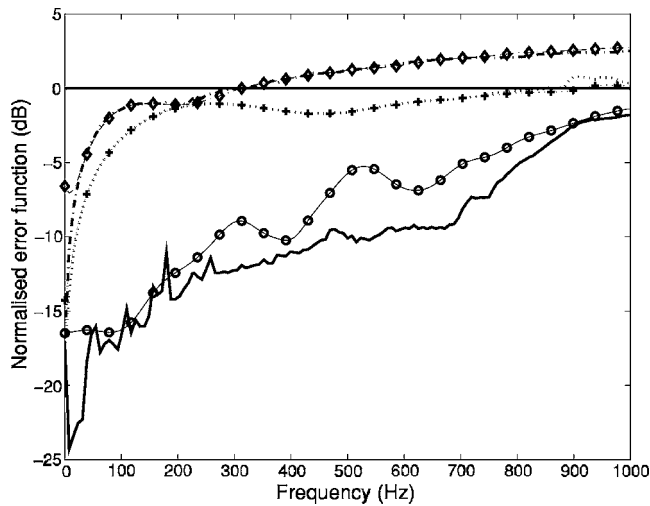
FIG. 7. Residual errors for the experimental synthesis of random excitations, those predicted from the measured plant responses (diffuse: bold line; plane wave: dotted line; TBL: dash-dotted line) and those experimentally synthesized (diffuse:circles; plane wave: crosses; TBL: diamonds).

array better matches the geometry of the microphone grid. Criteria are now formulated on the number of loudspeakers required per correlation length for an accurate experimental synthesis of the different kinds of random excitations.

B. Experimental synthesis of the random excitations

The statistics of an acoustic diffuse field, a grazing incident plane wave, and TBL wall-pressure fluctuations are experimentally synthesized with an array of 4×4 loudspeakers when optimally driven by the outputs of the AWG. For each kind of excitation, Fig. 7 shows the variations of the residual errors associated to either the experimental synthesis of the random pressures or those predicted from the measured plant response. It is found that the correlation patterns of a diffuse field and a grazing incident plane wave can be accurately reconstructed, respectively, up to about 1 kHz and 650 Hz whereas the spatial statistics of a TBL pressure field can only be synthesized up to about 200 Hz. These frequency ranges for the synthesis correspond to a −1 dB normalized residual error whereas a positive normalized residual error (in logarithmic scale) results in an erroneous reconstruction of the assumed correlation structures.

The results obtained for the experimental synthesis are in good agreement with those predicted from the measured transfer functions and discussed in Sec. III. It shows the robustness of the experimental reconstruction with respect to small changes that necessarily occur between the acoustic transfer functions first measured for the off-line calculation of the optimal drive signals and then equalized by the controller during the synthesis process. Such variations in the plant response are mostly due to inaccuracies in the positioning of the microphones over the surface of simulation as well as electronic and acoustic background noise.

The experimentally reconstructed correlation structures for a diffuse field and a grazing plane wave are shown respectively in Figs. 2 and 3 at frequencies that fall within and towards the upper frequency ranges of simulation. Figure 4 shows results for the experimental synthesis of a TBL exci-

TABLE I. Criteria on the number of acoustic sources required for the synthesis of a number of random pressure fields.

	Criteria on the residual error	Criteria on the mean-square error
Acoustic diffuse field[a]	1.9×1.5	2.7×2.2
Grazing plane wave[a]	2.8×2.3	3.4×2.7
TBL pressure field[b]	0.9×0.7	3.7×3.0

[a]Number of acoustic sources per unit acoustic wavelength.
[b]Number of acoustic sources per unit spanwise correlation length.

tation at frequencies above the frequency range of simulation. It can be verified that the statistical properties of the random fields are indeed accurately reproduced as long as the residual error is less than −1 dB. Beyond this limit, as a result of an insufficient number of acoustic sources, the amplitudes of the restored correlation structures start to deviate from the assumed statistical properties.

Empirical criteria are summarized in Table I on the number of sources required for an accurate reconstruction of the spatial correlation structures. They can be used as guidelines for further simulations with a denser array of loudspeakers. It can be seen that for all the excitations the residual error criteria are less stringent than those associated to the mean-square error: a −1 dB residual error corresponds to about a −5 dB mean-square error. Indeed, the peak values of the simulated correlation functions mostly contribute to the mean-square errors associated to the power spectral densities of the microphones outputs. Even though the peak values of the simulated correlation structures are underestimated and the −10 dB mean-square error criterion is not yet achieved, a reduced set of loudspeakers is still sufficient to reproduce an approximate correlation structure with the correct correlation area, which is a lower bound for the residual error [see Eq. (5)].

V. RESULTS ON THE PANEL RESPONSE

We now investigate the sensitivity of the panel vibroacoustic response to inaccuracies in the simulated excitations. A representative vibroacoustic model is developed in the Appendix to calculate the CSD matrices associated to the response of the randomly excited panel. This model is used to predict the approximate panel response induced by the random excitations synthesized in the loudspeakers array experiments. We only consider the simulation of an acoustic diffuse field and a TBL pressure field since the conclusions associated to the synthesis of a grazing plane wave are not significantly different from those obtained with a diffuse field, although occurring over a more limited frequency range.

A. Physical limitations

Assuming that the panel acoustic response is sufficiently well reproduced if the mean-square error J_p is less than −10 dB, it can be seen from Fig. 8 that accurate reproduction of the panel acoustic response can be achieved up to about 750 and 120 Hz when subject, respectively, to an acoustic diffuse field and a TBL pressure field. Moreover, a 4×4

C. Maury and T. Bravo: Experimental synthesis of random pressure fields
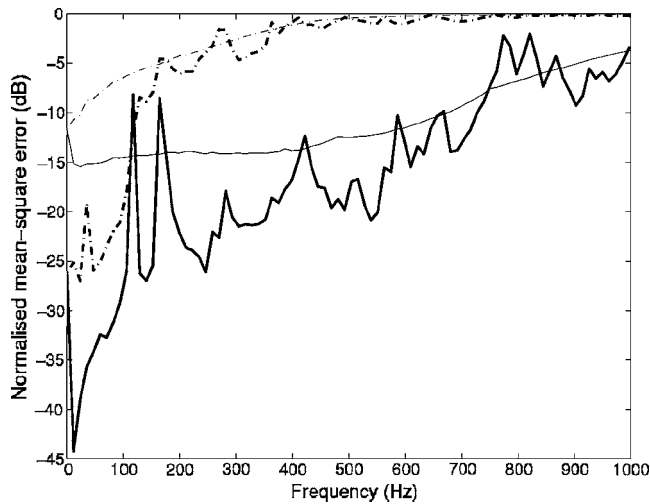
FIG. 8. The normalized mean-square error induced on the panel acoustic response when a diffuse field (thick solid) and a TBL (thick dash-dotted) are simulated together with the corresponding normalized mean square excitation error (respectively, thin solid and thin dash-dotted).
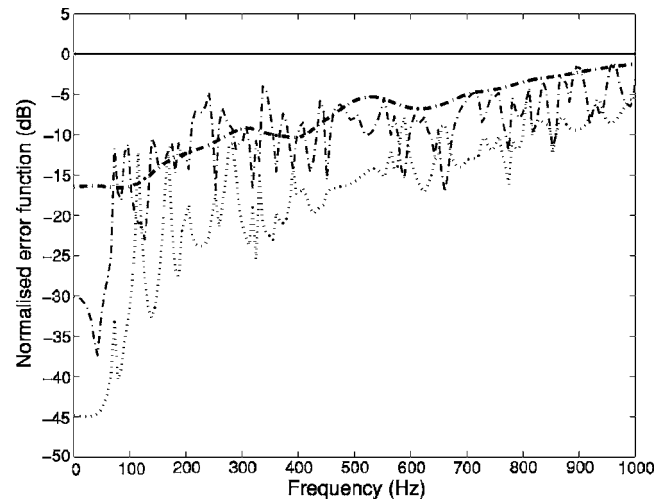
FIG. 9. Residual errors associated with the experimental synthesis of a diffuse field (dash-dotted thick), with the induced structural response (dash-dotted thin) and with the induced acoustic response (dotted).

array of suitably driven loudspeakers is able to achieve a rather small mean-square error associated to the panel acoustic response over these frequency ranges whereas little error reductions are achieved for the reproduction of the excitation fields over the same frequency bands. The reason for this is the wave number filtering effect of the panel acoustic response on the random excitation. Indeed, the sound pressure field radiated by the panel is much more spatially coherent than the approximate random excitation field and it can be reconstructed with a fewer number of independent components.

Also we note that the mean-square error induced on the panel response significantly rises at the panel resonances. Indeed, the optimal matrix of control filters only compensates for the acoustic transfer paths between the loudspeakers and the microphones over the simulation surface, but it does not equalize the transfer paths between the loudspeakers and the microphones on the radiating side of the panel, which takes into account the structure resonances. It is anticipated that the error associated to the direct simulation of the panel acoustic response would not exhibit such a resonant behavior, but on the other hand, it would require a preliminary modal analysis of the test panel to determine the optimum matrix of control filters.

B. Experimental results

An acoustic diffuse field is synthesized over the test panel and the residual errors as well as the correlations structures of both the excitation and the induced response are illustrated, respectively, in Figs. 9 and 10. Plots of the same nature are shown in Figs. 11 and 12 when associated to the reproduction of a TBL. For clarity, the correlation structures for the velocity responses in Figs. 10 and 12 have been plotted over a dense set of 26×32 panel elements, but calculations of the residual errors have only been made with 13×16 elemental surfaces.

Figures 9 and 11 indicate that, for both kinds of excitations and up to about 1 kHz, reasonable approximations to

the required spatial acoustic responses can already be achieved with a reduced set of 4×4 loudspeakers. Therefore, the restored correlation structures associated to the panel acoustic responses are generally less sensitive than the excitations to inaccuracies in the synthesis technique. It is also true for the reconstruction of the panel structural response (and near-field acoustic response) but to a lesser extent. We note that these results are less stringent than those based on the mean-square error reduction.

When the response is dominated by resonant modes, a larger residual error reduction is observed in the panel velocity response with respect to the excitation. In this case, the spatial coherence is governed by the resonant mode shapes of the panel and it is weakly sensitive to the approximate nature of the excitation. However, when the panel exhibits a nonresonant behavior, there is a decrease in the spatial coherence of the vibrational field, which therefore shows a greater dependency on imperfections in the simulated excitation. Despite this limitation, the velocity correlation associated to the nonresonant panel response induced at 990 Hz by a synthesized diffuse field is still well restored, with a residual error less than −6 dB (see Fig. 10).

In Fig. 12, the panel structural response to a TBL excitation is dominated at 587 Hz by the resonant mode $(6,1)$ which is weakly excited due to moderate coupling with the inaccurate TBL correlation pattern. This explains the slightly irregular shape observed for the assumed structural response. According to the criteria formulated in Table I, a minimum number of 12×12 loudspeakers would be required for an accurate synthesis of the TBL excitation up to this frequency and clearly, the loudspeakers array used in the experiment is not dense enough to reconstruct the required excitation. Despite this strong limitation, it can be seen that the panel velocity response is quite similar to the assumed one, as well as the acoustic response, with a residual error less than −9 dB. In this case, the resonant mode $(6,1)$ is highly excited and well couples with the inaccurate structure of the simulated TBL.

Theory   **Experiment, $\varepsilon_y^{exp} = -1.7$ dB**

$\varepsilon_v^{exp} = -6.4$ dB

$\varepsilon_p^{exp} = -6.6$ dB
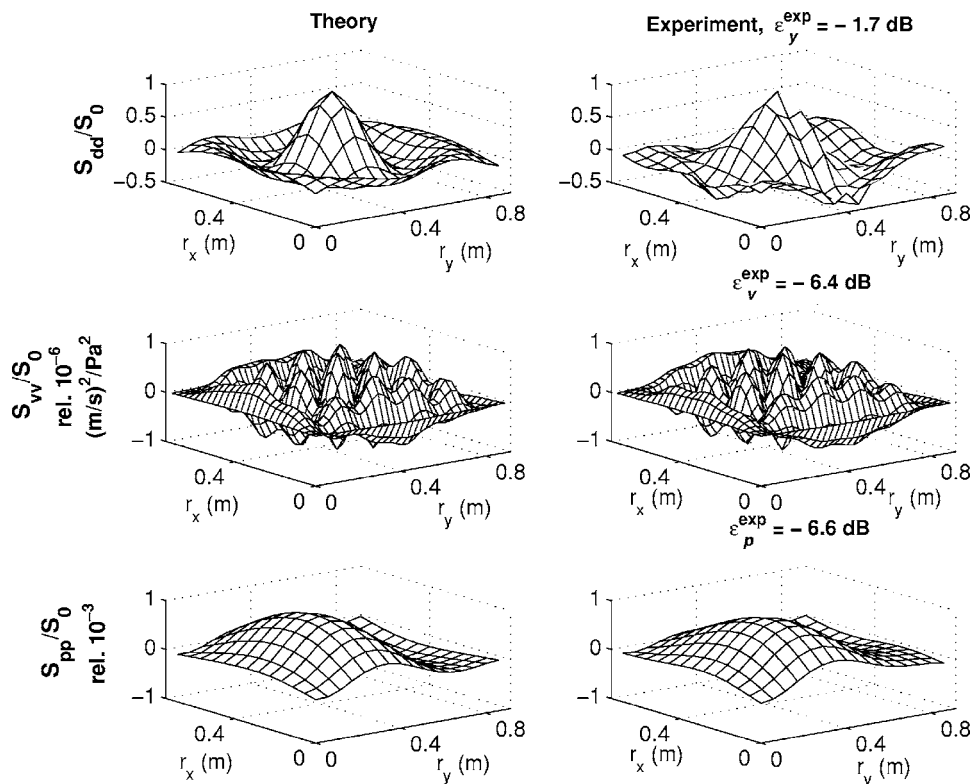
FIG. 10. Spatial correlation structures at 990 Hz associated with the excitation (top) and with the induced structural and acoustic responses (respectively, middle and bottom), when perfect reproduction of a diffuse field is assumed (left) and when it is experimentally synthesized with 4×4 loudspeakers (right).

In Fig. 13, the spectra of the panel kinetic energy are plotted for either a diffuse or a TBL excitation [see Eq. (A9)]. Figure 14 shows the corresponding sound power radiated by the panel [see Eq. (A10)]. In each case, the response spectra are calculated per unit point-power spectral density of the excitations and are therefore associated to the same mean-square wall-pressure value. We observe from the assumed responses that the panel is more efficiently excited by an acoustic diffuse field than by a TBL. At a given frequency, an acoustic diffuse field is correlated over an area greater than the area associated to a TBL excitation, and so it better couples with the panel modes than a TBL.[13]

When comparing the assumed and experimentally simu-

FIG. 11. Residual errors associated with the experimental reconstruction of the TBL pressure field (dash-dotted thick), with the induced structural response (dash-dotted thin) and with the induced acoustic response (dotted).

lated responses, it can be seen that the loudspeakers array is able to reconstruct the spectra induced by a diffuse field excitation up to about 1 kHz. However, the response spectra induced by the simulated TBL progressively deviate from the desired spectra beyond 300 Hz. This is due to the limited number of loudspeakers used in the experiment. In particular, the 4×4 array of loudspeakers cannot synthesize random fields correlated over an area lower than about 40% of the panel area. This lower frequency is reached at about 1 kHz for a synthesized diffuse field whereas it is already achieved at about 300 Hz for a simulated TBL [see Figs. 10 and 12]. Beyond these frequencies, the simulated excitations couple in a similar way with the panel modes and the response spectra induced by the reconstructed random fields tend to become very similar to each other.

VI. CONCLUSIONS

The present paper has shown the practical feasibility of synthesizing various types of random pressure fields with given statistical properties when using a reduced number of suitably driven near-field acoustic sources. The experimental technique based on the off-line calculation of the optimal drive signals and their subsequent synthesis by an arbitrary wave form generator has proved to be successful for reconstructing the spatial statistics of an acoustic diffuse field and a grazing incident plane wave, respectively, up to about 1 kHz and 650 Hz, both in terms of physical performance and robustness. However, it was found that the spatial statistics of TBL wall-pressure fields induced by high speed flows can be synthesized with an acceptable accuracy, only up to about 200 Hz, mainly due to an insufficient number of sources per unit correlation length. The spatial criteria that
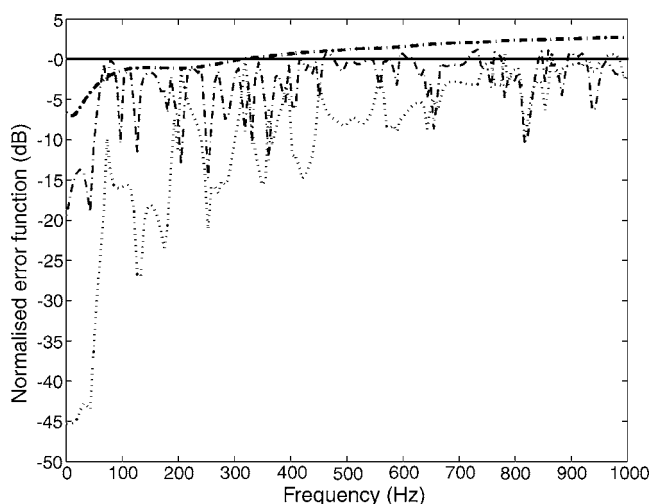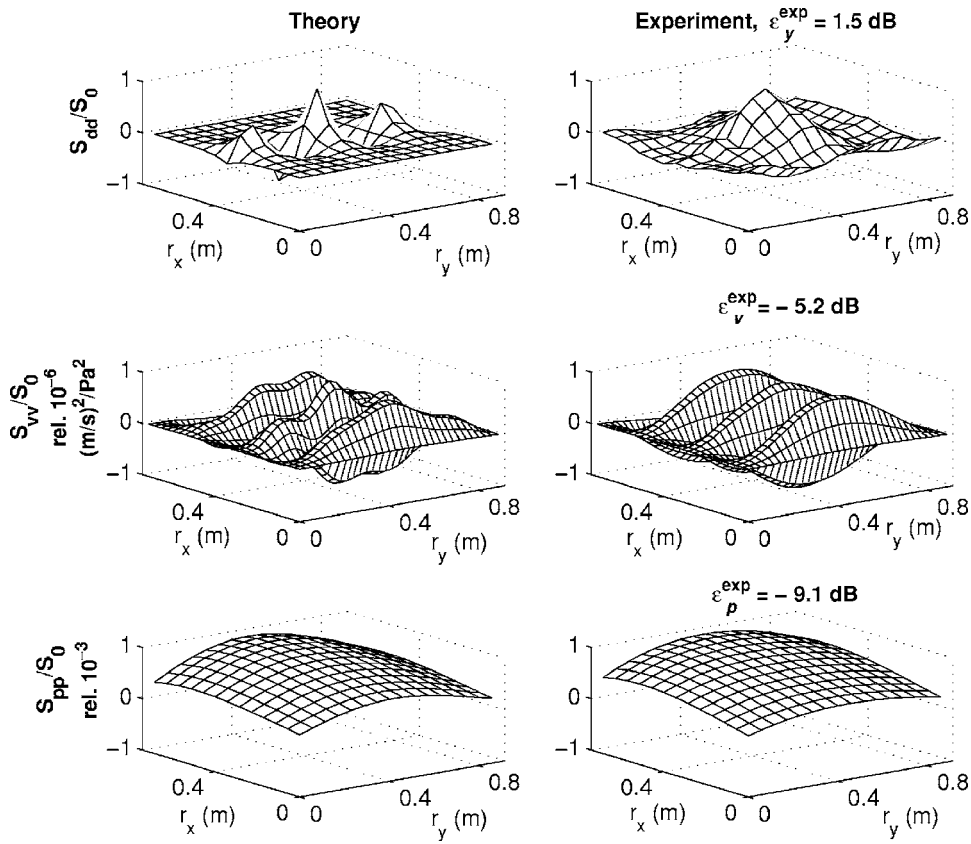
FIG. 12. Spatial correlation structures at 587 Hz associated with the excitation (top) and with the panel structural and acoustic responses (respectively, middle and bottom), when perfect reproduction of a TBL is assumed (left) and when it is experimentally synthesized with 4×4 loudspeakers (right).

have been formulated on the required number of loudspeakers are found to be less stringent than the mean-square error criteria which are related to the incident power of the signals synthesized at the microphones outputs.

From this latter energy-based criterion, it is found that the experimental synthesis of an acoustic diffuse field would require at most 2.7 acoustic sources per unit acoustic wavelength which is close to the theoretical lower bound of two sources per unit acoustic wavelength. An excellent reproduction of an acoustic diffuse field could thus be experimentally achieved up to 680 Hz in a semianechoic environment. Fur-

ther work should investigate how the loudspeakers array would perform in a reverberant transmission suite below the Schroeder frequency. The simulation performances may be worsened by the nonminimum phase behavior of the plant transfer matrix to equalize.

At most 3.4 sources per unit trace of the acoustic wavelength are required to simulate experimentally the spatial correlation structure of an incident plane wave of random phase and amplitude. It corresponds to an upper frequency limit of 550 Hz which covers the frequency range above which acoustic progressive wave tubes are able to simulate wideband wall-pressure fluctuations equivalent, in power
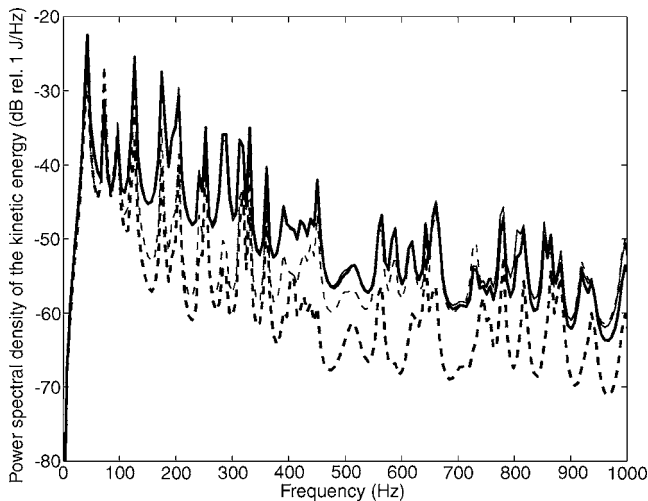
FIG. 13. Kinetic energy of the panel when subject to either a diffuse field (assumed: bold; experimentally simulated: thin) or a TBL (assumed: dashed bold; experimentally simulated: dashed thin).
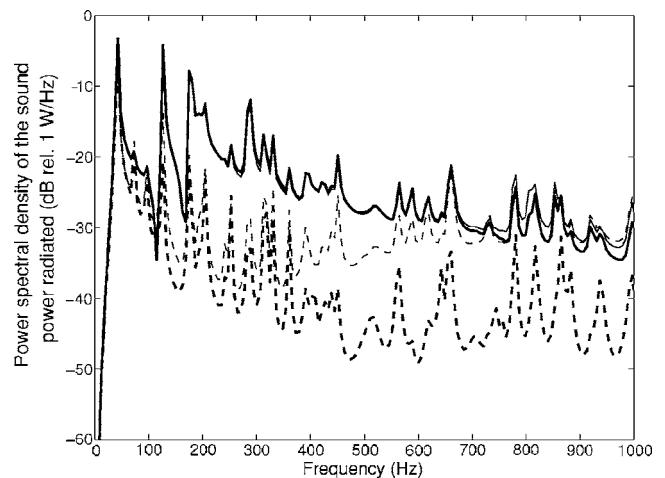
FIG. 14. Sound power radiated by the panel when subject to either a diffuse field (assumed: bold; experimentally simulated: thin) or a TBL (assumed: dashed bold; experimentally simulated: dashed thin).

spectral density, to jet-induced or propeller-induced excitations. However, very stringent high intensity as well as thermal issues have to be addressed before the experimental rig could be used for fatigue testing purposes.

The simulation of a TBL pressure field with an array of 4×4 uniformly distributed loudspeakers appears to be limited to the very low frequency range and to high speed flows due to the rapid decay of the spanwise correlation length as the frequency increases. Experimental simulation of the TBL correlation structure would require about 3.7 acoustic sources per unit spanwise correlation length. A greater number of loudspeakers as well as a denser distribution along the spanwise direction would contribute to enlarge the frequency range of simulation. However, given that the surface of simulation would be representative of an aircraft skin panel typically of dimensions $0.23 \text{ m} \times 0.25 \text{ m}$, this would require loudspeakers of smaller size with necessarily limited performances at low frequencies. A solution would be to directly simulate the vibroacoustic response of a tensioned skin panel using the experimental rig designed in this work. Good performances are expected since the small size of the test section together with the tensioning effect should enhance the wave number filtering capabilities of the panel and therefore limit the number of sources required.

Most of the physical limitations observed for the simulation of each of the three pressure fields could be overcome through the direct simulation of the panel vibroacoustic response. We have found that imperfections in the reproduction technique have little influence on the panel acoustic response, especially at or near resonances. From a physical point of view, direct simulation could *a priori* be feasible within the frequency bandwidth for which both resonant and coincident modes dominate the panel vibroacoustic response. In particular, the direct reconstruction of the panel response induced by a TBL should be feasible over the hydrodynamic coincidence area, i.e., over the frequency range for which the spacewise variations of the resonant modes are mostly coincident with those of the forcing field. Future work could be directed towards the experimental implementation of such simulation techniques.

ACKNOWLEDGMENTS

APPENDIX: THE VIBROACOUSTIC MODEL

The statistics of the panel velocity response are described by a spectral density function, $S_{\nu\nu}$, between two points $\mathbf{M}$ and $\mathbf{M}'$ located over the test panel of surface Σ. $S_{\nu\nu}$ is expressed in terms of the spectral density of the excitation simulated over the panel, S_{yy}, as

$$S_{\nu\nu}(\mathbf{M},\mathbf{M}';\omega) = B_\omega^{\mathbf{M},\mathbf{M}'}(G_\nu, G_\nu), \qquad (A1)$$

where B_ω is a bilinear form defined as

$$B_\omega^{\mathbf{M},\mathbf{M}'}(u,v) = \int\int_\Sigma \int\int_\Sigma u(\mathbf{M},\mathbf{M}'';\omega)S_{yy}(\mathbf{M}'' - \mathbf{M}''';\omega)$$
$$\times v^*(\mathbf{M}',\mathbf{M}''';\omega)d^2\mathbf{M}''d^2\mathbf{M}''', \qquad (A2)$$

and $G_\nu(\mathbf{M},\mathbf{M}'';\omega)$ is the panel velocity response at a point $\mathbf{M}$ due to a unit point force excitation acting at the point $\mathbf{M}''$, or transfer mobility function. The transfer function G_ν is decomposed as a series of the normal modes of the test panel. This model has been updated in order to be representative of the panel kinetic energy up to 1 kHz, i.e., over the frequency range of simulation.[1] The power spectral density of the panel's kinetic energy is defined as

$$S_{E_c}(\omega) = \frac{\rho h}{2} \int\int_\Sigma S_{\nu\nu}(\mathbf{M},\mathbf{M};\omega)d^2\mathbf{M}, \qquad (A3)$$

where ρ is the panel density, h_p is the panel thickness, and $S_{\nu\nu}$ is obtained from (A1).

An integral representation of the sound pressure field radiated by the panel enables to express the statistical properties, S_{pp}, associated to the acoustic response between two points $\mathbf{P}$ and $\mathbf{P}'$ at two microphones locations, in terms of the spectral density of the excitation reproduced over the panel, S_{yy}, as

$$S_{pp}(\mathbf{P},\mathbf{P}';\omega) = B_\omega^{\mathbf{P},\mathbf{P}'}(G_p, G_p), \qquad (A4)$$

with

$$G_p(\mathbf{P},\mathbf{M}'';\omega) = j\rho_0\omega \int\int_\Sigma G_\nu(\mathbf{M},\mathbf{M}'';\omega)$$
$$\times \frac{e^{-jk|\mathbf{P}-\mathbf{M}|}}{2\pi|\mathbf{P}-\mathbf{M}|}d^2\mathbf{M}. \qquad (A5)$$

G_p is the pressure field radiated at a point $\mathbf{M}$ when the panel is excited by a harmonic point force located at a point $\mathbf{M}''$ on Σ. ρ_0 is the density of the air and k is the acoustic wave number.

The power spectral density of the sound power radiated by the panel is defined as

$$S_\pi(\omega) = \frac{1}{2}\Re\left[\int\int_\Sigma S_{p\nu}(\mathbf{M},\mathbf{M};\omega)d^2\mathbf{M}\right], \qquad (A6)$$

where the cross-spectral density function $S_{p\nu}$ is given by $S_{p\nu}(\mathbf{P},\mathbf{M}';\omega) = B_\omega^{\mathbf{P},\mathbf{M}'}(G_p, G_\nu)$.

The discretized version of the integrals (A1) and (A4) enables to express the CSD matrices of the panel vibrating and acoustic responses, respectively, $\mathbf{S}_{\nu\nu}$ and $\mathbf{S}_{pp}$, in terms of the CSD matrix of the excitation, $\mathbf{S}_{yy}$, as follows:

$$\mathbf{S}_{\nu\nu} = \mathbf{G}_\nu \mathbf{S}_{yy} \mathbf{G}_\nu^H, \qquad (A7)$$

$$\mathbf{S}_{pp} = \mathbf{G}_p \mathbf{S}_{yy} \mathbf{G}_p^H, \qquad (A8)$$

where $\mathbf{G}_\nu$ and $\mathbf{G}_p$ are the plant response matrices whose elements are, respectively, generated by G_ν and G_p. The CSD matrix (A8) is calculated between the far-field pressures ra-

diated by the panel and measured over a plane grid of 13 × 16 microphones positioned one acoustic wavelength apart from the radiating surface.

After discretizing the integrals (A3) and (A6), one obtains the following expressions for the spectra of both the kinetic energy and the sound power radiated by the panel:

$$S_{E_c}(\omega) = \frac{\rho h_p s}{2} \text{Tr}[\mathbf{S}_{\nu\nu}], \tag{A9}$$

$$S_{\pi}(\omega) = \text{Tr}[\mathbf{S}_{\nu\nu}\mathbf{R}], \tag{A10}$$

where s is the area of each panel element and $\mathbf{R}$ is the radiation resistance matrix with entries $\mathbf{R}_{\{i,j\}} = (\rho_0 \omega s^2 k/4\pi)$ sinc(kr_{ij}), r_{ij} is the distance between the midpoints of the ith and jth panel elements.

[1] T. Bravo and C. Maury, "The experimental synthesis of random pressure fields: Methodology," J. Acoust. Soc. Am. 120(5) (in press).

[2] C. J. Dodds, "The laboratory simulation of vehicle service stress," J. Eng. Ind., 391–398 (1974).

[3] C. J. Dodds and J. D. Robson, "The description of road surface roughness," J. Sound Vib. 31, 175–183 (1973).

[4] A. Zerva and V. Zervas, "Spatial variation of seismic ground motions: An overview," Appl. Mech. Rev. 55, 271–297 (2002).

[5] A. Steinwolf, "Shaker simulation of random vibrations with a high kurtosis value," Journal of the Institute of Environmental Sciences 15, 33–43 (1997).

[6] M. Shinozuka, "Simulation of multivariate and multidimensional random processes," J. Acoust. Soc. Am. 49(2), 357–367 (1971).

[7] C. T. Hugin, A. P. Payne, D. C. Tomlinson, and K. H. Heron, "Simulation of turbulent boundary layer excitation by point array forcing experiment," Report on EU Framework V Project G4RD-CT-2000-00223, Environmental Noise Associated with Turbulent Boundary Layer Excitation: EN-ABLE, QinetiQ Ltd, Farnborough, UK (2002).

[8] C. Maury, S. J. Elliott, and P. Gardonio, "Turbulent boundary layer simulation with an array of loudspeakers," AIAA J. 42, 706–713 (2004).

[9] S. J. Elliott, C. Maury, and P. Gardonio, "The synthesis of spatially correlated random pressure fields," J. Acoust. Soc. Am. 117, 1186–1201 (2005).

[10] T. Bravo and C. Maury, "Experimental reproduction of random pressure fields in laboratory conditions," J. Acoust. Soc. Am. 117(2), 2547 (2005).

[11] C. Maury and S. J. Elliott, "Analytic solutions of the radiation modes problem and the active control of sound power," Proc. R. Soc. London, Ser. A 461, 55–78 (2005).

[12] H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I (Wiley, New York, 2001), Chap. 3, pp. 166–237.

[13] C. Maury, P. Gardonio, and S. J. Elliott, "A wavenumber approach to modelling the response of a randomly excited panel, Part I: General Theory," J. Sound Vib. 252, 83–113 (2002).

Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays

Heinz Teutsch[a)] and Walter Kellermann[b)]

Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany

This paper is concerned with the problem of detecting and localizing multiple wideband acoustic sources by applying the notion of wavefield decomposition using circular microphone arrays optionally mounted into cylindrical baffles. The decomposed wavefield representation is used to serve as a basis for so-called modal array signal processing algorithms, which have the significant advantage over classical array signal processing algorithms that they inherently support multiple wideband acoustic sources. A rigorous derivation of modal array signal processing algorithms for source detection and localization, as well as performance evaluations, by means of measurements using an actual real-time capable implementation are presented. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2346089]

I. INTRODUCTION

In this paper, the problem of extracting information contained in an acoustic space, such as the wavefields or parameters describing the acoustic sources, is addressed. Since acoustic wavefields are processes spread out in space and time, it follows quite naturally that these tasks are predominantly performed by evaluating the signals captured by a number of spatially distinct microphones, i.e., microphone arrays. A standard and widely applied vehicle for evaluating the microphone array signals is built upon classical array signal processing techniques.[1,2] Hereby, the term "classical" is used to denote signal processing algorithms that are applied directly to the individual microphones comprising the array. In contrast, the algorithms presented in this paper are applied to signals that are obtained by transforming the microphone signals into a domain defined by the eigensolutions of the acoustic wave equation.

This paper deals with the extraction of the desired source signals from the observed wavefield by means of beamforming and beamsteering techniques.[3] Also, the wavefield parameters, i.e., the number and location of possibly multiple simultaneously active wideband acoustic sources, are of interest for, e.g., acoustic surveillance and for applications based on beamforming as a preprocessing step. While the localization of a *single* desired acoustic source has been the subject of research for several decades—mainly by estimating the time difference of arrival (TDOA) between microphone pairs[4–6]—the problem of estimating the number and positions of *multiple* acoustic sources using nonheuristic methods has only been investigated very recently.[7–12] Motivated by this observation, this paper introduces algorithms that fundamentally support the detection and location of multiple wideband acoustic sources.

The algorithms presented here are not solely based on the foundations of classical array signal processing but also on the principles of wave propagation and wave scattering offered by classical acoustics. Algorithms for parameter estimation solely based on "classical array signal processing" often suffer from a narrowband assumption underlying the signal model. This assumption limits these algorithms' applicability when wideband signals, such as speech, are present in the wavefield under observation. The popular ESPRIT algorithm for source localization tasks is an example of a method that relies on a narrowband assumption.[13] It will be subsequently shown that by also considering methods described by "classical acoustics," algorithms can be derived that are inherently wideband; hence, allowing for more general application. The possibility to formulate parameter estimation algorithms that inherently allow for wideband source models is a previously unknown result of the fusion between classical array signal processing and methods of classical acoustics.

Moreover, it will be shown that acoustic source localization, based on these novel methods, has the potential to unambiguously detect and localize multiple simultaneously active wideband sources in the array's full field of view, i.e. all possible azimuthal angles of arrival.

In this paper, the physics of wave propagation in two-dimensional space is considered for application to localization and detection of multiple wideband acoustic sources. Two-dimensional wavefields can be used as reasonable models for propagating acoustic sound fields in closed rooms, where ceiling and floor reflections are sufficiently attenuated. A natural way of analyzing a two-dimensional wavefield is to decompose it into an orthogonal set of eigenfunctions of the acoustic wave equation in cylindrical coordinates. It will be subsequently shown that this decomposition into so-called "eigenbeams" (EBs) (Ref. 14) can be obtained by circular apertures mounted into a rigid cylindrical baffle, as indicated in Fig. 1.

[a)]Now with Avaya Labs, 233 Mount Airy Rd., Basking Ridge, NJ 07920; electronic address: teutsch@research.avayalabs.com
[b)]Electronic address: wk@lnt.de
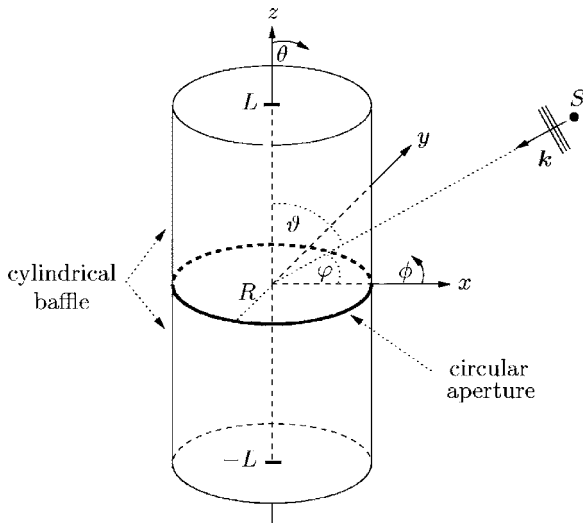
FIG. 1. Geometric model.

The remainder of this paper is organized as follows. Section II briefly reviews acoustic wavefields created by cylindrical radiators, and the wavefield's modification due to the presence of cylindrical scatterers. The wavefield decomposition process utilizing continuous circular apertures, as well as circular microphone arrays, is outlined in Sec. III. Waveform and parameter estimation techniques employing the EBs, as obtained from the wavefield decomposition process, are developed in Sec. IV. In particular, a novel formulation of the popular ESPRIT algorithm, which allows for the localization of multiple wideband acoustic sources, is presented in Sec. IV B. An acoustic source detection procedure based on EBs is derived in Sec. IV C. Performance evaluations of both algorithms are presented in Sec. V and conclusions are drawn in Sec. VI.

II. ACOUSTIC WAVEFIELDS IN CYLINDRICAL COORDINATES

Assuming linearity of the transmitting medium and steady-state conditions governing all particle interactions, the homogeneous Helmholtz equation describing acoustic wavefields with respect to pressure P in cylindrical coordinates, (r, ϕ, z), can be written as,[15]

$$\nabla^2 P(r, \phi, z) + k^2 P(r, \phi, z) = 0, \qquad (1)$$

where the Laplace operator is given by

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2}. \qquad (2)$$

Note that the frequency-dependence in terms of the wave number $k = 2\pi f/c$, where f is the frequency and c is the speed of sound, has been omitted for notational convenience.

For *interior* problems, i.e., where all sources are located outside of the cylindrical surface under consideration, the general steady-state solution of the Helmholtz equation, Eq. (1) is[16]

$$P(r, \phi, z) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{jn\phi}\int_{-\infty}^{\infty} A_n(k_z)e^{jk_z z}J_n(k_r r)dk_z, \qquad (3)$$

where

$$k_r = \sqrt{k^2 - k_z^2} \qquad (4)$$

are the radial components, and k_z are the z-dependent components of the wave vector $\boldsymbol{k}$, and $k = \|\boldsymbol{k}\|$. $J_n(\cdot)$ denotes the nth-order Bessel function. Note that the integral in Eq. (3) is a Fourier integral. Therefore, the quantity $A_n(k_z)$ can be interpreted as Fourier coefficients. The geometric arrangement for an interior problem is shown in Fig. 1, where the source "S" is located outside the cylindrical boundary at $r = R$.

For *exterior* problems, i.e., where all sources are located inside the cylindrical surface under consideration, the general steady-state solution of the Helmholtz equation in cylindrical coordinates is[16]

$$P(r, \phi, z) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} e^{jn\phi}\int_{-\infty}^{\infty} B_n(k_z)e^{jk_z z}H_n(k_r r)dk_z, \qquad (5)$$

where $H_n(\cdot) \triangleq H_n^{(1)}(\cdot)$ denotes the nth-order Hankel function of the first kind and $B_n(k_z)$ are the appropriate Fourier coefficients. A vibrating cylindrical structure, such as the one indicated in Fig. 1, can be considered as an exterior problem.

For simplicity, only normal wave incidence with respect to the z axis is considered in the following discussions, i.e., $\vartheta = \pi/2$, cf. Fig. 1. Assuming further, for now, that $\varphi = 0$, then the pressure due to a single far-field source for $\theta = \pi/2$ and $L \to \infty$ can be written as[17]

$$P_\infty(kr, \phi) = \sum_{n=-\infty}^{\infty} j^n C_n(kr)e^{jn\phi}. \qquad (6)$$

Depending on the applicable boundary conditions,[16,18] it follows that

$$C_n(kr) \triangleq \begin{cases} J_n(kr), & \text{unbaffled} \\ J_n(kr) - \dfrac{J_n'(kR)}{H_n'(kR)}H_n(kr), & \text{baffled-1} \\ J_n(kr) - \dfrac{J_n(kR)}{H_n(kR)}H_n(kr), & \text{baffled-2,} \end{cases} \qquad (7)$$

where the prime associated with the Bessel and Hankel functions denotes the derivative with respect to the argument. Hereby, "unbaffled" means that the circular aperture is not mounted into a cylindrical baffle. "Baffled-1" denotes a circular aperture mounted into an acoustically rigid cylindrical baffle of infinite length, i.e., a rigid scatterer. Here, the acoustic impedance of the baffle is much higher than the acoustic impedance of the transmitting medium resulting in the observation that a large portion of the acoustic energy is reflected without a phase change. "Baffled-2" denotes a circular aperture mounted into a acoustically soft, or pressure-release, cylindrical baffle. In this case, the acoustic impedance of the baffle is much smaller than the acoustic impedance of the transmitting medium resulting in the observation that a large part of the acoustic energy is reflected with a phase change of π.

An important result to be noted here is that the pressure on the *surface* of an infinite-length cylindrical scatterer at $z=0$ follows with $r=R$ from Eq. (6) as

$$P_\infty(kR,\phi) = \sum_{n=-\infty}^{\infty} j^n C_n(kR)e^{jn\phi}. \qquad (8)$$

In the following, only rigid cylindrical baffles are considered, i.e., the "baffled-1" case. The extension of the results to soft cylindrical baffles, i.e., the "baffled-2" case, is straightforward.

Of course, infinite-length cylindrical scatterers are only of theoretical interest. In practice, finite-length cylindrical scatterers need to be considered. Neglecting the boundary conditions at the endcaps, it can be shown that the pressure on the surface of a finite-length rigid cylindrical scatterer, as depicted in Fig. 1 at $z=0$ due to normal plane wave incidence, can be expressed as[17]

$$P_L(kR,\phi) = \sum_{n=-\infty}^{\infty} j^n \left[J_n(kR) - \left(\frac{kL}{\pi} J_n'(kR) \right. \right.$$
$$\left. \left. \times \int_{-\infty}^{\infty} \frac{H_n(\sqrt{k^2-k_z^2}R)\mathrm{sinc}(k_z L)}{\sqrt{k^2-k_z^2}H_n'(\sqrt{k^2-k_z^2}R)}dk_z \right) \right] e^{jn\phi}, \qquad (9)$$

where $\mathrm{sinc}(k_z L) = \sin(k_z L)/(k_z L)$. Note that since[19]

$$\lim_{L\to\infty} \frac{L}{\pi} \mathrm{sinc}(k_z L) = \delta(k_z), \qquad (10)$$

where $\delta(\cdot)$ denotes the Dirac distribution, it follows immediately that

$$\lim_{L\to\infty} P_L(kR,\phi) = P_\infty(kR,\phi), \qquad (11)$$

in the "baffled-1" case.

III. WAVEFIELD DECOMPOSITION

This section details the wavefield decomposition process obtained by circular apertures. For simplicity, the theoretical construct of a continuous circular aperture is considered first. Then, sampled circular apertures, i.e., circular microphone arrays are utilized for the decomposition process. Both apertures are optionally mounted into a rigid cylindrical baffle as indicated in Fig. 1.

A. Continuous circular apertures

A quantity of considerable interest for the following discussion is the response of a circular aperture due to a plane-wave impinging from $(\vartheta = \pi/2, \varphi)$ emitted by a source "S" which is located in the far-field of the aperture, cf. Fig. 1. Both infinite-length rigid cylindrical scatterers, as well as finite-length cylindrical scatterers, are discussed next.

1. Infinite-length cylindrical scatterer

The total response of a circular aperture mounted into an infinite-length rigid cylindrical scatterer can be obtained by applying a spatial Fourier transform[16] to the total incident pressure measured at the circular aperture, Eq. (8), after weighting the aperture by a frequency-dependent function $w(k,\phi)$. Physically, this means that all contributions of the plane wave on the aperture are weighted and integrated to yield the total response of the aperture. Therefore, the response can be expressed as

$$F_m(kR,\varphi) \triangleq \frac{1}{2\pi} \int_0^{2\pi} w(k,\phi)P_\infty(kR,\phi-\varphi)e^{-jm\phi}d\phi, \qquad (12)$$

and can be interpreted as the mth-Fourier series coefficient of the weighted pressure distribution along the aperture. The weighting function can be chosen arbitrarily, although here it will be set to $w(k,\phi)=1$, resulting in an omnidirectionally sensitive aperture. Note that, when spatial sampling of the continuous aperture is considered in Sec. III B, this choice corresponds to employing omnidirectional microphones. With Eq. (8), Eq. (12) becomes

$$F_m(kR,\varphi) = \frac{1}{2\pi} \int_0^{2\pi} \sum_{n=-\infty}^{\infty} j^n C_n(kR)e^{jn(\phi-\varphi)}e^{-jm\phi}d\phi. \qquad (13)$$

By applying the orthonormality property of the exponential function, Eq. (13) can be simplified to finally yield

$$F_m(kR,\varphi) = j^m C_m(kR)e^{-jm\varphi}, \qquad (14)$$

and with $n \triangleq m$,

$$F_n(kR,\varphi) = j^n C_n(kR)e^{-jn\varphi}. \qquad (15)$$

Note that the expression in Eq. (15) can be found as atomic components in the solutions of the Helmholtz equation, cf. Eq. (3) and Eq. (5). Therefore, Eq. (15) represents an eigensolution of the Helmholtz equation in cylindrical coordinates. The symbol $F_n(kR,\varphi)$ is sometimes also referred to as a mode, circular harmonic, or EB of order n.[14] It is obvious from Eq. (15) that these EBs exhibit a response whose frequency-dependent component, $C_n(kR)$, is decoupled from the direction-dependent component, $e^{-jn\varphi}$. Figure 2 shows the normalized frequency-dependent so-called *modal* magnitude response of the first four eigenbeams, $20\log|F_n(kR,\varphi)|$, $n=0,1,2$, and 3. Several properties of the magnitude response are of particular interest here. First, for $n>0$, all eigenbeams exhibit a highpasslike characteristic, with a slope of about $6n$ dB/octave up to the normalized frequency $kR \approx n$. This, in turn, means that a decomposed wavefield, at very small kR, only has the zero-order (omnidirectional) component of considerable strength. As kR increases, more and more eigenbeams are gaining in strength. Second, for an unbaffled aperture, all eigenbeams show periodic zeros in the magnitude response which correspond to the zeros of the respective Bessel functions $J_n(\cdot)$. As a consequence, signals that carry components in the vicinity of the zeros in the magnitude responses cannot be completely identified. By mounting the aperture into a rigid cylindrical baffle, the zeros in the magnitude response disappear, cf. Fig. 2(b). This can be explained by the fact that the Bessel function and the second term in Eq. (7) do not share any common zeros. An additional very interesting property of baffled apertures can be deduced by closely examining and comparing Figs.
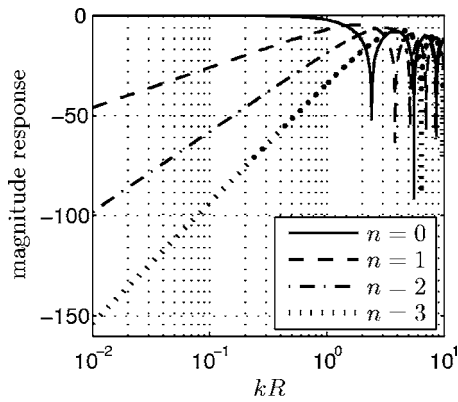
(a) unbaffled aperture
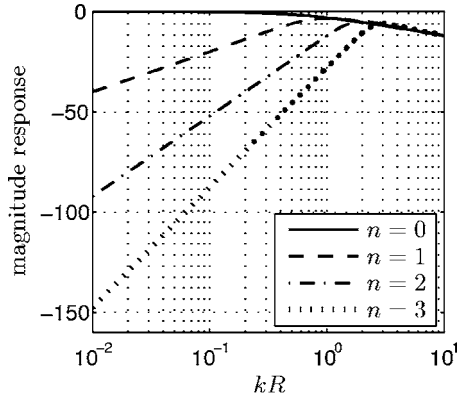
(b) infinite-length rigid cylindrical aperture

FIG. 2. Modal magnitude response of a continuous circular aperture.

2(a) and 2(b). For small kR and $n>0$, the magnitude response due to the presence of a rigid baffle obviously lies a few dB above the response of the unbaffled circular aperture. This effect can be interpreted as a *virtual increase of the modal aperture* due to the presence of the scatterer. This effect is examined further in the following. Expressing the relative magnitude response between the response of a circular aperture, mounted into a rigid cylindrical baffle and an unbaffled circular aperture, gives for $\varphi=0$

$$\Delta F_n(kR) \triangleq \frac{J_n(kR) - J_n'(kR)H_n(kR)/H_n'(kR)}{J_n(kR)}$$

$$= \frac{2j}{\pi kR J_n(kR)H_n'(kR)}, \tag{16}$$

where—in the last step—a Wronskian relationship[20] has been used. Since the virtual aperture increase is only effective for small kR, the small argument approximations $(kR \ll n)^{16}$

$$J_n(kR) \approx \frac{(kR)^n}{2^n n!} \tag{17}$$

$$H_n'(kR) \approx \frac{jn!}{\pi \epsilon_n} \left(\frac{2}{kR} \right)^{n+1}, \tag{18}$$

where
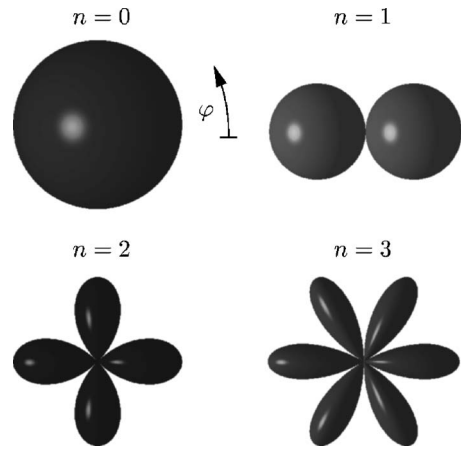
FIG. 3. Normalized azimuthal response of the EBs with respect to φ for $n=0,1,2,$ and 3.

$$\epsilon_n = \begin{cases} 1, & n=0 \\ 2, & n>0 \end{cases} \tag{19}$$

applied to Eq. (16) yields

$$\Delta F_n(kR) \approx \epsilon_n. \tag{20}$$

It therefore follows from Eq. (20) that for all $n>0$, the rigid scatterer increases the response of the circular aperture by $20 \log_{10}(2)=6.02$ dB. For quantifying the modal aperture increase, one has to equate the magnitude response of the aperture with the larger radius, $\widetilde{R}$, to the one of "normal" radius, R, multiplied by ϵ_n, i.e., $|J_n(k\widetilde{R})| = \epsilon_n |J_n(kR)|$. With Eq. (17), it follows that the effective modal aperture for $kR \ll n$ increases by a factor of $\Delta R \triangleq \widetilde{R}/R = \sqrt[n]{\epsilon_n}$.

Figure 3 shows the azimuth-dependent component of the EBs, i.e., $e^{-jn\varphi}$. For visualization, $|\text{Re}\{e^{-jn\varphi}\}| = |\cos(n\varphi)|$ is plotted. As can be seen, the eigenbeams correspond to multipoles, i.e., monopole, dipole, quadrupole, etc. Note that these multipoles are mutually *orthonormal*.

2. Finite-length cylindrical scatterer

Applying the same line of arguments as for the infinite-length rigid scatterer, the EBs associated with a finite-length rigid scatterer follow from Eq. (9) as

$$G_n(kR, \varphi) = j^n \left[J_n(kR) - \frac{kL}{\pi} J_n'(kR) \right.$$

$$\left. \times \int_{-\infty}^{\infty} \frac{H_n(\sqrt{k^2 - k_z^2}R)\text{sinc}(k_z L)}{\sqrt{k^2 - k_z^2} H_n'(\sqrt{k^2 - k_z^2}R)} dk_z \right] e^{-jn\varphi}. \tag{21}$$

It can be observed that the azimuth-dependent response is not a function of the scatterer's length. The frequency-dependent response, however, undergoes moderate modifications as the length of the scatterer changes. As an example, Fig. 4 shows the modal magnitude response of a continuous circular aperture mounted into a finite-length rigid cylindrical baffle for $L/R=\{0, 0.1, 1.4, \infty\}$. Note that $L/R=0$ corresponds to an unbaffled aperture, while $L/R=\infty$ corresponds to an aperture mounted into an infinite-length rigid

(a) $n = 0$

(b) $n = 2$
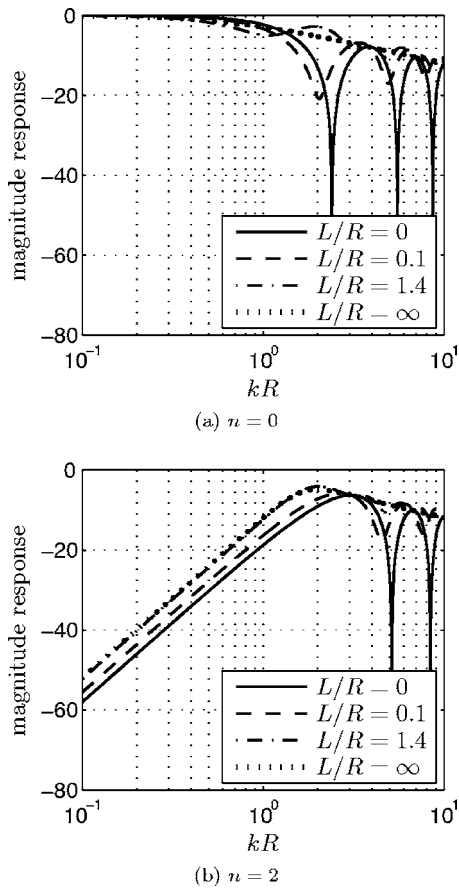
FIG. 4. Modal magnitude response of a continuous circular aperture mounted into a finite-length rigid cylindrical baffle.

cylindrical scatterer, cf. Fig. 2(b). Note also that $L/R = 1.4$ corresponds to the geometry of the experimental system used for performance evaluations, see Fig. 10 in Sec. V. It can be concluded that even a relatively short cylindrical scatterer is capable of combating the zeros in the frequency response resulting from unbaffled circular apertures. Therefore, continuing further derivations based on an infinite-length rigid cylindrical scatterer seems justified.

B. Circularly symmetric microphone arrays

Up until this point, the discussion exclusively dealt with continuous circular apertures. For real-world applications, however, the continuous aperture needs to be sampled at discrete points in space, i.e., by microphones. In this section, it is assumed that this sampling is performed by ideal omnidirectional microphones. It is further assumed that one wishes to decompose an impinging wavefield into EBs up to a finite order $\mathcal{N}$. Note that in Figs. 2 and 3, $\mathcal{N} = 3$. Following the same reasoning as for the sampling theorem of one-dimensional time-domain signals,[21] a minimum of

$$M \geq 2\mathcal{N}, \tag{22}$$

microphones are necessary for this task.[1] Equality in Eq. (22) denotes the special case of *critical sampling*, which can only be employed when it is ensured that the sensors do not sample the zero crossings of the cosine function, which

would result in a zero-valued output. This sampling theorem can easily be verified by noting that when considering EBs up to order $\mathcal{N}$ eigenbeams, $2\mathcal{N}+1$ Fourier coefficients of the *truncated* Fourier series describing the total wavefield at the circular aperture [cf. Eqs. (8) and (9), where the symbol "∞" is to be replaced by "$\mathcal{N}$"] need to be identified by at least $2\mathcal{N}+1$ samples.

It can be shown[22] that the sampled eigenbeams can be expressed as

$$F_n^s(kR,\varphi) = j^n C_n(kR)e^{-jn\varphi} + \sum_{q=1}^{\infty} j^{Mq-n} C_{Mq-n}(kR)e^{j(Mq-n)\varphi}$$

$$+ \sum_{q=1}^{\infty} j^{Mq+n} C_{Mq+n}(kR)e^{-j(Mq+n)\varphi}, \tag{23}$$

where the superscript "s" indicates the sampling operation applied to the continuous circular aperture. The first term in Eq. (23) is equivalent to the eigenbeam obtained by continuous circular apertures, cf. Eq. (15). The remaining terms are residuals due to the sampling operation. Sampling a continuous circular aperture, therefore, results not only in an EB of order n but also, inseparably superimposed, in EBs of order $n \geq M/2$. In analogy to the error occurring due to sampling of one-dimensional time-domain signals, this distortion is denoted as *modal aliasing*. However, there is one significant difference between one-dimensional time-domain signals and two- or three-dimensional space-time signals. Time-domain signals can be bandlimited in an effective way before application of any type of digital signal processing. This can be achieved by applying analog low-pass filters of sufficiently high order. Therefore, a strict rule can be formulated which quantifies how aliasing can be avoided. This rule is the well-known Nyquist criterion. It is very difficult, if not impossible, to perform effective band-limitation for space-time signals before the application of digital signal processing algorithms, since high-order spatial lowpass filters do not seem to exist. Hence, there is no strict rule guaranteeing an error-free sampling operation. As a consequence, modal aliasing cannot be entirely avoided. However, by judiciously choosing the number of microphones, M, and the maximum spatial frequency, kR, spatial aliasing can be controlled. As an example, Fig. 5 shows the eigenbeam of order $n = 2$ obtained by a continuous baffled circular aperture and a baffled circular microphone array comprising $M = 10$ sensors. Note that Fig. 5(a) can be seen as a combination of Figs. 3 and 2(b) for order $n = 2$. Modal aliasing is clearly visible in Fig. 5(b) for kR greater than about 5, but virtually absent for $kR < 4$.

Evaluation of Eq. (23) shows that the effect of modal aliasing increases with increasing order and decreases with increasing number of sensors. Since no strict aliasing criterion can be specified, a design procedure may include a maximum amount of error to be introduced due to modal aliasing by selecting an appropriate amount of sensors and frequency range of operation. The frequency range of operation has to be restricted since the incoming wavefield is, in general, not bandlimited.

H. Teutsch and W. Kellermann: Wavefield decomposition and applications

(a) $F_2(kR, \varphi)$

(b) $F_2^s(kR, \varphi)$, $M = 10$
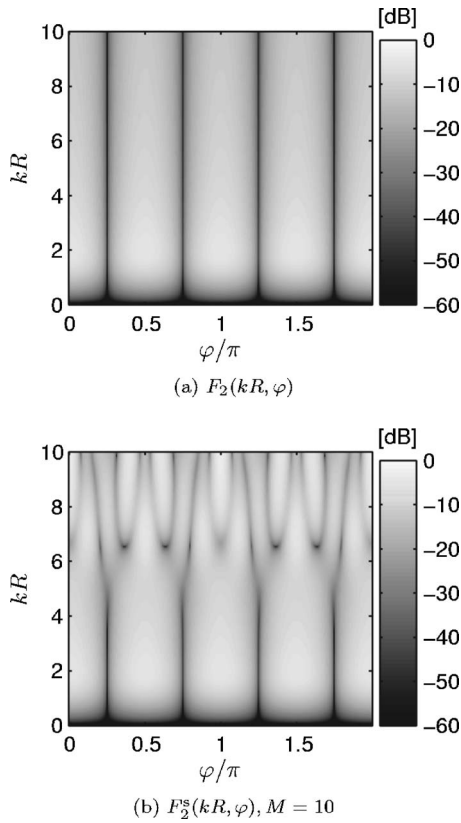
FIG. 5. EB of order $n=2$ obtained by a continuous baffled circular aperture and a circular microphone array.

IV. APPLICATIONS OF MODAL ARRAY SIGNAL PROCESSING

In this section, the EBs as derived in Sec. III are used as a basis for a novel concept in array signal processing applied to waveform and parameter estimation tasks, here denoted as *modal* array signal processing. In particular, Sec. IV A briefly discusses the application of modal array signal processing to beamforming, which results in the so-called eigen-beamformer (EBF).[14,23] Sections IV B and IV C detail the application of EBs to acoustic source localization and source detection problems, respectively.

A. Beamforming: EBF

Beamformers using EBs obtained by spherical apertures and arrays have been introduced previously by using the notion of a modal beamformer.[14,23] This section shows how the paradigm of eigenbeamforming can be applied to continuous circular apertures. It is assumed that a planar wavefield has been decomposed into eigenbeams in the sense of Sec. III A 1 and that all processing is applied in the modal domain. In the context of beamforming, these EBs, or multipoles, can be used as building blocks for more complicated beampatterns as shown next.

In general, any square-integrable function on a circle can be expanded into a series of circular harmonics, or eigenbeams. In particular, any desired beampattern, $F^d(kR, \phi)$, with look direction—or rotation angle—$\varphi_s \in \{0 \ldots 2\pi\}$, can be written as a Fourier transform pair as,
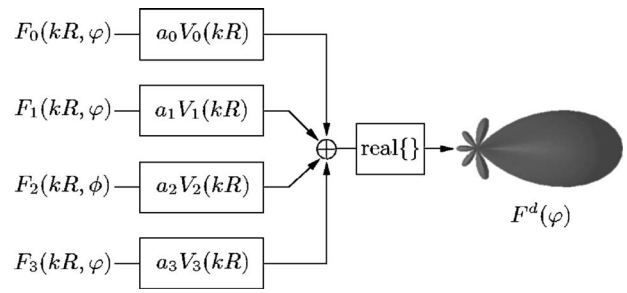
FIG. 6. Example of an EBF for $\mathcal{N}=3$.

$$F^d(kR, \phi, \varphi_s) = \sum_{n=-\infty}^{\infty} F_n^d(kR)e^{jn(\phi-\varphi_s)}, \qquad (24)$$

$$F_n^d(kR) = \frac{1}{2\pi} \int_0^{2\pi} F^d(kR, \phi, \varphi_s)e^{-jn(\phi-\varphi_s)}. \qquad (25)$$

Truncating the series to the $\mathcal{N}$th harmonic, *any* beampattern of order up to and including $\mathcal{N}$ can be synthesized. Letting $\phi=\varphi$ and considering a finite number of EBs, i.e., $\mathcal{N}+1$, for real-valued $F^d(kR, \varphi, \varphi_s)$, Eq. (24) can also be written as

$$F^d(kR, \varphi, \varphi_s) = \sum_{n=0}^{\mathcal{N}} \epsilon_n F_n^d(kR)\cos[n(\varphi - \varphi_s)], \qquad (26)$$

where ϵ_n is defined in Eq. (19). Assume one desires to design a frequency-independent third-order hypercardioid pattern for $\varphi_s=0$, i.e.,

$$F^d(\varphi) = \sum_{n=0}^{3} a_n\cos(n\varphi), \qquad (27)$$

where for cylindrically isotropic wavefields[24]

$$\boldsymbol{a} \triangleq [a_0\ a_1\ a_2\ a_3]^T = \frac{1}{7}[1\ 2\ 2\ 2]^T. \qquad (28)$$

Then, since the desired pattern is real valued, it can be readily verified—by comparing Eqs. (26) and (27)—that the EBF structure in Fig. 6 is obtained, where

$$V_n(kR) \triangleq F_n^d(kR) = \frac{1}{j^n C_n(kR)}, \quad n = 0,1,2,3. \qquad (29)$$

It can be deduced that $V_n(kR)$ basically defines a filter for equalizing the magnitude response of the individual eigenbeams, $F_n(kR, \varphi)$, which in turn results in a frequency-independent beampattern, $F^d(\varphi)$. Note that the resulting pattern can be easily rotated by considering a rotation $\varphi_s \neq 0$ in Eq. (26). Note also that, in this section, only continuous circular apertures were considered. An extension to circular microphone arrays is straightforward.

B. Source localization: EB-ESPRIT

Most source localization algorithms for wideband acoustic sources are based on an evaluation of the TDOA between microphone pairs.[4–6] The problem here is that the underlying signal model only allows for a single source to be active at any given time in space.

In this section, a method is derived for *multiple* wideband acoustic source localization based on a popular subspace-based algorithm, namely ESPRIT.[1,13] However, the classical ESPRIT algorithm applied to individual sensor signals, i.e., processing done in *sensor space*, requires the signals to be emitted by narrowband sources. It will be subsequently shown that by applying the notion of EBs to the classical ESPRIT algorithm, an inherently frequency-independent localization method, i.e., EB-ESPRIT, can be devised which does not suffer from this limitation. As a result, EB-ESPRIT performs all processing in so-called *modal space*.

Let $\mathcal{L}$ far-field sources impinge on the circular aperture shown in Fig. 1. Prior to the application of the ESPRIT algorithm, the frequency dependence of the individual EBs must be compensated for, so that

$$\tilde{F}_n(\varphi) \triangleq V_n(kR)F_n(kR,\varphi) \approx e^{-jn\varphi}, \tag{30}$$

where the approximation indicates the assumption that $F_n^s(kR,\varphi) \approx F_n(kR,\varphi)$, i.e., the error due to the sampling operation is assumed to be sufficiently small. Then, the modal array response, i.e., the time-domain output of the discrete spatial Fourier transform (DFT) along the aperture, at time-instance t can be written as

$$y(t) = As(t) + n(t), \tag{31}$$

where

$$A = [F(\varphi_1)| \ldots |F(\varphi_{\mathcal{L}})], \tag{32}$$

is a *frequency-independent* modal array matrix with

$$F(\varphi_\ell) = [\tilde{F}_{-\mathcal{N}}(\varphi_\ell), \ldots, \tilde{F}_{\mathcal{N}}(\varphi_\ell)]^T, \quad \ell = 1,2,\ldots,\mathcal{L}, \tag{33}$$

and where the source vector is defined as $s(t) = [s_1(t),\ldots,s_{\mathcal{L}}(t)]^T n(t)$ denotes the noise vector and $(\cdot)^T$ denotes transposition. The noise field is assumed to be independent of the source signals. Therefore, a total of $Q = 2\mathcal{N}+1$ eigenbeams can be used for further processing.

Dropping the explicit time dependence in the following for notational convenience, the covariance matrix of the modal array response is

$$R_{yy} = E\{yy^H\}, \tag{34}$$

where $E\{\cdot\}$ denotes the expectation operator and $(\cdot)^H$ is the Hermitian operator. Assuming equal power of and statistical orthogonality between the Gaussian and spatially white-noise contributions to each eigenbeam, Eq. (34) can be expressed as

$$R_{yy} = AR_{ss}A^H + \sigma^2\tilde{I}, \tag{35}$$

where R_{ss} is the signal covariance matrix, $\tilde{I}$ denotes the identity matrix, and σ^2 is the noise variance. Let $\xi_l, l = 1,2,\ldots,Q$, denote the lth eigenvector of the covariance matrix R_{yy}. Then, it can be shown[1] that the $\mathcal{L}$ largest eigenvalues, corresponding to the $\mathcal{L}$ signal sources, form the so-called signal subspace E_S as,
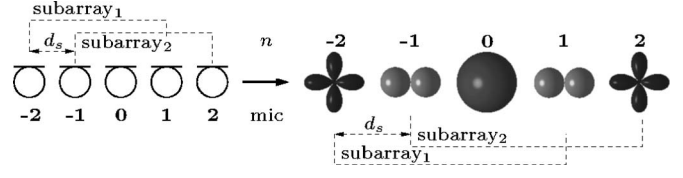
FIG. 7. Transition from sensor space to modal space for $\mathcal{N}=2$ and displacement $d_s=1$.

$$E_S \triangleq [\xi_1|\xi_2| \ldots |\xi_{\mathcal{L}}], \tag{36}$$

and that they are linear combinations of the modal array vectors of A. This observation is of central importance for all subspace-based source localization algorithms. Based on the definition of the signal subspace, it can be shown that it is possible to localize up to $\mathcal{L}=Q-1$ simultaneously impinging sources,[1,17] leaving only one eigenvector for the noise subspace with dimension one.

Now, the standard sensor-space ESPRIT[13] can be directly applied to the "modal-space" after replacing the notion of an individual microphone of a standard linear microphone array by an individual eigenbeam of order n, see Fig. 7. It is assumed that the first element in the original sensor space (modal space) array is the first element in the first subarray (modal subarray), and that the (d_s+1)th element in the original array is the first element in the second subarray. Note that d_s, therefore, does *not* correspond to a physical shift in either array configuration. The motivation for this transition can be justified by realizing that the modal array matrix A in Eq. (32) is Vandermonde-type, just like the array matrix of a linear array.[1] A significant difference, however, is the fact that A is now *frequency independent*.

Following the sensor-space ESPRIT algorithm,[13] two overlapping modal subarrays of length $\tilde{M}$ and displacement d_s, cf. Fig. 7, are chosen, whose subarray (modal subarray) matrices satisfy the invariance relation:

$$A_2 = A_1\Phi, \tag{37}$$

where

$$\Phi = \text{diag}\{e^{-jd_s\cdot\varphi_1}, \ldots, e^{-jd_s\cdot\varphi_{\mathcal{L}}}\}, \tag{38}$$

and

$$A_1 = [K|0] A, \tag{39}$$

$$A_2 = [0|K] A, \tag{40}$$

where K is an $\tilde{M} \times \tilde{M}$ identity matrix and 0 is a $\tilde{M} \times d_s$ zero matrix, e.g., 4×1 in Fig. 7. Since the columns of A span the signal subspace, E_S, it holds that

$$E_S = AT, \tag{41}$$

where T is a nonsingular matrix. Therefore, the subspaces of the two subarrays can be defined as

$$E_{S_1} = A_1T, \tag{42}$$

and

$$E_{S_2} = A_2T = A_1\Phi T. \tag{43}$$

By combining Eqs. (42) and (43), one obtains

$$E_{S_2} = E_{S_1}\mathbf{\Psi}, \tag{44}$$

where

$$\mathbf{\Psi} = T^{-1}\mathbf{\Phi}T. \tag{45}$$

$\mathbf{\Psi}$ can then be obtained from Eq. (44) by applying a standard least-squares or total least-squares solver.[25] By realizing that the eigenvalues of $\mathbf{\Psi}$ are the diagonal elements of $\mathbf{\Phi}$, the locations of the sources can be estimated. Note, that in a real system, all equalities in Eqs. (41)–(43) must be replaced by approximate equalities since they must be estimated from the observed covariance matrix, $\hat{R}_{yy}$, which is subject to measurement errors.

Note also, although not shown here, that for a single source an algorithm can be formulated that does not require the frequency compensation used in Eq. (30).[17] This fact has the advantage that uncorrelated noise does not get amplified at low frequencies by low-passlike equalization filters exhibiting a slope of about $6n$ dB/octave, cf. Fig 2.

The fact that individual microphones can be replaced by individual EBs is of central importance in the derivation of EB-ESPRIT. It can be shown[17] that this idea can also be applied to other subspace-based source localization algorithms, such as unitary-ESPRIT,[26] which is basically a real-valued formulation of ESPRIT, and the popular MUSIC algorithm.[27] By applying a temporal Fourier transform to all signals introduced in this section, a frequency-domain formulation of the EB-ESPRIT algorithm can be readily obtained. The performance of EB-ESPRIT in the time-domain is discussed in Sec. V. Note that though offering an improved computational efficiency, EB-ESPRIT—in the frequency-domain—yields equivalent performance compared to the time-domain implementation.

C. Source detection: EB-DETECT

All subspace-based source localization techniques require the number of currently active sources as *a priori* information, cf. Eq. (36). This section describes a detection algorithm for circular apertures which is based on the wavefield decomposition process outlined in Sec. III. In this context, it is important to note that the decomposition of a wavefield into EBs is performed with respect to the aperture's perimeter. It can therefore be observed that, after applying this transformation, one obtains a single virtual receiver exhibiting multiple spatial selectivity characteristics, i.e., the EBs of order n, simultaneously. As mentioned earlier, $n=0$ corresponds to an omnidirectional spatial characteristic, $n=1$ corresponds to a dipole spatial characteristic, and so on.

First, the unbaffled circular aperture is considered. Then, it is explained how to modify the algorithm when baffled circular apertures are employed. It is assumed that the effect of modal aliasing can be neglected due to proper array design, i.e., $F_n^s(kR,\varphi) \approx F_n(kR,\varphi)$.

Let $\mathcal{L}$ plane waves, with unity amplitude, impinge on an unbaffled circular aperture. Then, due to the superposition principle, it follows in modal space that

$$P(kR,\phi,\boldsymbol{\varphi}) = \sum_{\ell=1}^{\mathcal{L}} e^{jkR\cos(\phi-\varphi_\ell)} = \sum_{\ell=1}^{\mathcal{L}}\sum_{n=-\infty}^{\infty} j^n J_n(kR)e^{jn(\phi-\varphi_\ell)}, \tag{46}$$

where $\boldsymbol{\varphi}$ denotes the angular dependency due to the directions of arrival (DOAs) of all $\mathcal{L}$ impinging plane waves, $\boldsymbol{\varphi} = [\varphi_1,\varphi_2,\ldots,\varphi_{\mathcal{L}}]^T$. As a result, it now follows for the EBs corresponding to *multiple* plane waves that, cf. Eq. (15),

$$F_n(kR,\boldsymbol{\varphi}) = \sum_{\ell=1}^{\mathcal{L}} j^n J_n(kR)e^{-jn\varphi_\ell}. \tag{47}$$

As a first step for deriving an algorithm that estimates the number of acoustic wideband sources, a wavefield *synthesis* operation that superimposes the limited number of eigenbeams is considered. This synthesis operation can be seen as an order-limited inverse spatial Fourier series expansion with respect to $\phi=\phi_0$ cf. Eq. (24), which gives with Eq. (47)

$$P(kR,\phi_0,\boldsymbol{\varphi}) = \sum_{n=-\mathcal{N}}^{\mathcal{N}} F_n(kR,\boldsymbol{\varphi})e^{jn\phi_0} + \mathcal{E}_{\text{trunc}}(kR,\phi_0,\boldsymbol{\varphi}), \tag{48}$$

where

$$\mathcal{E}_{\text{trunc}}(kR,\phi_0,\boldsymbol{\varphi}) = \sum_{|n|>\mathcal{N}} F_n(kR,\boldsymbol{\varphi})e^{jn\phi_0} \tag{49}$$

is the error term due to the truncation of the infinite number of eigenbeams to a finite number.

For the sake of simplicity, the truncation error, $\mathcal{E}_{\text{trunc}}(kR,\phi_0,\boldsymbol{\varphi})$ is assumed to be sufficiently small in the frequency range of interest here. Then, Eq. (48) with Eq. (46) and $\phi_0=0$ can be expressed as

$$P(kR,0,\boldsymbol{\varphi}) = \sum_{\ell=1}^{\mathcal{L}} e^{jkR\cos\varphi_\ell} \approx \sum_{n=-\mathcal{N}}^{\mathcal{N}} F_n(kR,\boldsymbol{\varphi}). \tag{50}$$

The transformation of Eq. (50) into the time domain gives the *modal impulse response*

$$p(t,\boldsymbol{\tau}) = \sum_{\ell=1}^{\mathcal{L}} \delta(t-\tau_\ell), \tag{51}$$

where the symbol $\boldsymbol{\tau}$ combines the delays of all $\mathcal{L}$ impinging plane waves, $\boldsymbol{\tau} = [\tau_1,\tau_2,\ldots,\tau_{\mathcal{L}}]^T$ and, where

$$\tau_\ell = \frac{R}{c}\cos\varphi_\ell, \quad \ell = 1,2,\ldots,\mathcal{L}, \tag{52}$$

is a delay caused by the ℓth-plane wave with respect to $\phi_0=0$ on the aperture.

A rule for the system detection problem for wave fronts, propagating in free field can now be derived by observing that the number of Dirac impulses in Eq. (51), corresponds to the number of sources in the impinging wavefield.

Since the targeted application of EB-DETECT involves natural speech signals recorded in real acoustic environments, it follows quite naturally that the actual detection algorithm should be based on an adaptive identification of the
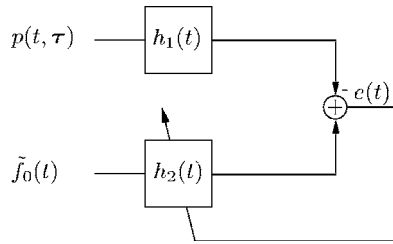
FIG. 8. Equivalent system identification problem.

modal impulse response, $p(t, \boldsymbol{\tau})$, from a block of observed data. In order to transform this task into a classical system identification problem,[28] a reference signal is required. This signal should be independent of the direction with respect to the incoming plane waves as well as of the number of sources. A suitable choice is the zeroth-order EB since it exhibits an omnidirectional spatial characteristic. This signal can be found in Eq. (47) for $n=0$. For the purpose of system identification, the frequency dependence of the zeroth-order harmonic needs to be equalized. The modified reference signal is, therefore,

$$\widetilde{F}_0(kR) \triangleq \frac{F_0(kR)}{J_0(kR)} = \mathcal{L}, \tag{53}$$

which in the time domain equals

$$\widetilde{f}_0(t) = \mathcal{L} \cdot \delta(t). \tag{54}$$

A schematic of the resulting system identification problem is shown in Fig. 8, where the filter $h_1(t)$ is simply a delay that compensates for the delay introduced by the adaptive filter $h_2(t)$, which should be identical to a delayed version of $p(t, \boldsymbol{\tau})$ after convergence.

In practice, this system identification procedure can be performed by applying any adaptive filtering technique such as least mean-squares-based or recursive least-squares-based algorithms.[28]

Figure 9(a) depicts the discretized modal impulse responses $p(t=\nu T, \boldsymbol{\tau})$, where ν is an integer and T is the sampling interval, and $\widetilde{f}_0(t=\nu T)$ for a single bandlimited plane-wave incident on a continuous aperture in free field for $\varphi = \varphi_1 = \pi/6$. The parameters are chosen as $R=0.04$ m, $\mathcal{N}=5$, $f_s=1/T=48$ kHz, and $c=340$ m/s, where f_s denotes the sampling frequency. The sensor signals were bandlimited to $f_{max}=7$ kHz, which corresponds to $(kR)_{max} \approx 5.2$.

Also shown are two vertical lines that represent the maximum delays $\tau_{max} = \pm R/c$, cf. Eq. (52), that need to be considered in this detection scheme. This observation is of some importance since it is expected that moderately reverberant environments only have a limited negative impact on the estimation of the modal impulse response in this very narrow section of the impulse response. Therefore, even though the signal model is based on a free-field model, moderate reverberation should not seriously affect the performance of the algorithm.

In Fig. 9(b), two simultaneously impinging plane waves are considered. The parameters chosen were the same ones that were used to produce Fig. 9(a); the difference being that $\boldsymbol{\varphi} = [\pi/6, 2\pi/3]^T$, i.e., $\varphi_1 = \pi/6$ and $\varphi_2 = 2\pi/3$, indicating si-
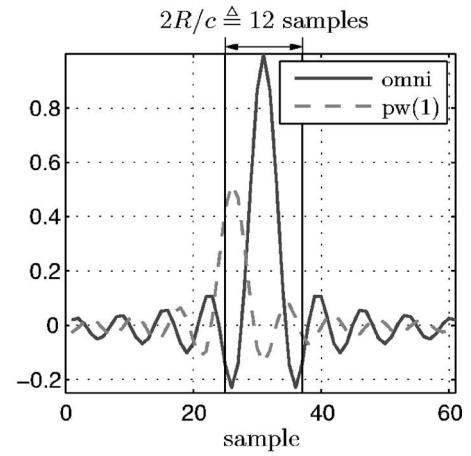
(a) single plane-wave, $\varphi = \pi/6$
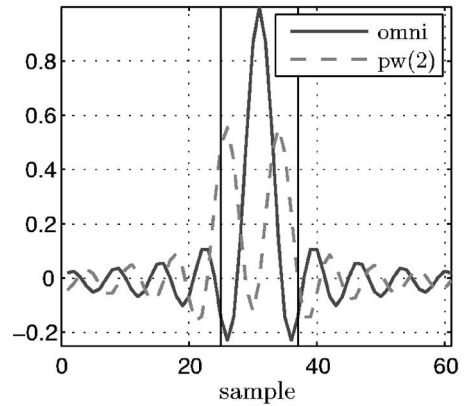
(b) two plane-waves, $\varphi_1 = \pi/6, \varphi_2 = 2\pi/3$

FIG. 9. Normalized modal impulse responses of the omnidirectional component, omni, and the plane wave component due to $\mathcal{L}$ plane waves, pw($\mathcal{L}$).

multaneously arriving wave fronts. Two distinct peaks in the modal impulse response appear, which correspond to the two sources.

However, as can be deduced from the cosine term in Eq. (52), the two sources cannot be located arbitrarily along the circle. For instance, a plane wave impinging from $\varphi_1 = \pi/6$ cannot be distinguished from a second plane wave impinging from $\varphi_2 = 5\pi/6$, since they cause an identical delay in Eq. (52). This ambiguity problem can be alleviated by also performing the wavefield-synthesis operation in Eq. (48) with respect to $\phi_0 \neq 0$, e.g., $\phi_0 = \pi/2$, and use both detection results for an overall estimate. Any value $\phi_0 \neq 0$ is equivalent to rotating the individual EBs by ϕ_0; thereby resolving the ambiguity problem.

Note that the method presented above cannot be directly applied to baffled apertures without modifications. A wavefield due to a superposition of $\mathcal{L}$ plane waves on a circular aperture mounted into a rigid cylindrical baffle is, cf. Eq. (8)

$$P(kR, \phi, \boldsymbol{\varphi}) = \sum_{\ell=1}^{\mathcal{L}} \sum_{n=-\infty}^{\infty} j^n \left[J_n(kR) \right.$$
$$\left. - \frac{J_n'(kR)H_n(kR)}{H_n'(kR)} \right] e^{jn(\phi-\varphi_\ell)}. \tag{55}$$

Applying the decomposition into EBs as described in Sec. III, yields

$$F_n(kR, \varphi) = \sum_{\ell=1}^{\mathcal{L}} j^n \left[J_n(kR) - \frac{J_n'(kR) H_n(kR)}{H_n'(kR)} \right] e^{-jn\varphi_\ell}. \quad (56)$$

A finite sum of these eigenbeams does not yield an estimate for the superposition of plane waves at the perimeter of the aperture. In contrast to unbaffled circular apertures, the EBs derived from baffled apertures not only represent a wavefield containing plane waves, but also an additional component corresponding to the scattered wave on the surface of the rigid cylinder. Therefore, the superimposed eigenbeams do not correspond to individual peaks in the time domain. Rather, additional peaks occur due to the scattered wavefield. This fact turns the detection task into a very difficult problem. One possibility to overcome this problem is to "undo" the wave scattering effect by a filtering operation. Note that the scattered wavefield only appears in the frequency-dependent component of the eigenbeams, whereas the angular-dependent component remains unaffected. By defining the modified EB of a baffled circular aperture as

$$\widetilde{F}_n(kR, \varphi) \triangleq \frac{F_n(kR, \varphi)}{\left[J_n(kR) - \dfrac{J_n'(kR) H_n(kR)}{H_n'(kR)} \right]} J_n(kR), \quad (57)$$

and choosing

$$\widetilde{F}_0(kR, \varphi) = \frac{F_0(kR, \varphi)}{\left[J_0(kR) - \dfrac{J_0'(kR) H_0(kR)}{H_0'(kR)} \right]} = \mathcal{L}, \quad (58)$$

the same algorithm as presented above for unbaffled circular apertures can then be directly transferred to the baffled circular aperture scenario. Note that the term in the denominator of Eqs. (57) and (58) defines a filter in the frequency domain equalizing the modal frequency response of an nth-order eigenbeam.

Note that by restricting oneself to the employment of *unbaffled* circular microphone arrays, the task of source detection as described in this section can also be accomplished by simply evaluating the relative time differences of plane-wave arrival between two microphones along the circular aperture.[6] The detection scheme based on two individual microphone fails, however, in the case of *baffled* circular microphone arrays since the effect of scattering depends on the plane waves' DOA. Therefore, this effect cannot be compensated for in sensor space by a digital filtering operation, while compensation is possible in modal space.

A few words on resolution capacity are in order, i.e., the question on how close the plane waves can be located along the circle with respect to each other. As can be seen in Fig. 9 and Eq. (52), the maximum delay that needs to be taken into consideration for the detection algorithm depends on the radius of the aperture as well as on the sampling frequency. In the example shown above, where $R=0.04$ m, $c=340$ m/s, and $f_s=48$ kHz, the maximum delay with respect to the peak corresponding to the omnidirectional term is only about ± 6 samples. This, of course, limits the minimum angular distance between the two plane waves, as the two peaks merge into one and can, therefore, not be resolved if the spatial separation is small. For a real system, it is very difficult, if
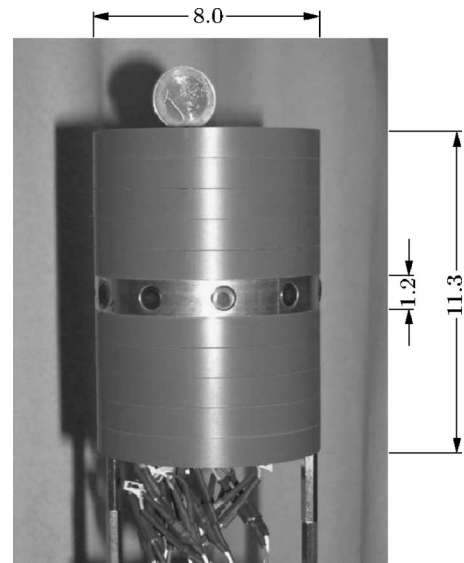
FIG. 10. Ten-sensor circular microphone array mounted into a rigid cylindrical baffle (dimensions in centimeters).

not impossible, to formulate criteria of general validity on the minimum spacing between the two sources since the accuracy of the identification of the modal impulse responses, depends on the sampling frequency, the frequency range of operation, which is limited due to the effects of wavefield truncation, aperture sampling, and the position of the two sources due to the properties of the cosine function. Note that resolution capacity also largely depends on the properties of the measured acoustic field itself.

V. SYSTEM EVALUATION

The algorithms derived in the preceding sections have been validated by both simulations as well as by measurements performed with an actual circular microphone array, with radius $R=0.04$ m, mounted into a finite-length rigid cylindrical baffle, see Fig. 10. A one Euro coin is placed on top of the array to stress its compact size. Ten calibrated omnidirectional 0.25 in. electret capsules were utilized.

Due to space limitations, the performance evaluations by means of simulations are omitted here. The interested reader is referred to Ref. 17 for a detailed description of the simulation results. In this paper, only performance evaluations obtained by measurements in real acoustic environments are presented in order to stress the fact that the algorithms as described in the previous sections are applicable to real-world scenarios.

The first set of measurements pertains to the array's ability to extract the EBs from a planar wavefield, cf. Sec. III. For this purpose, the array was placed into a low-reverberation chamber. Impulse responses were measured from a fixed loudspeaker position to the array which was constantly rotated by using a stepper motor. The wavefield was then decomposed by applying a spatial DFT to the recorded wavefield along the circular microphone array. As an example, the thus decomposed EB of order $n=2$ due to a sampled circular aperture is reproduced in Fig. 11. It can be directly compared with the simulated second-order EB pre-
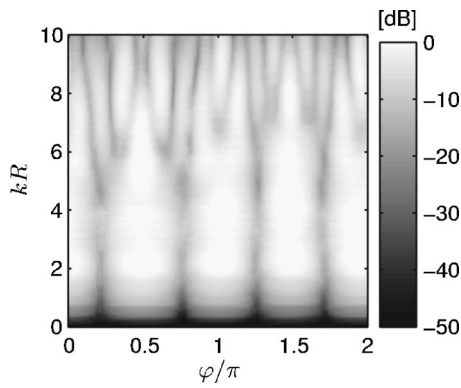
FIG. 11. Measured EB of order $n=2$.

(a) mean estimation error

(b) estimator variance

FIG. 13. Measured DOA estimation performance of EB-ESPRIT for $\mathcal{L}=2$.

sented in Sec. III B, see Fig. 5(b). Striking similarities suggest satisfactory agreement with the theory and that the realized finite-length cylindrical baffle is a sufficient approximation of the infinite-length cylindrical baffle considered in Sec. III B.

EB-ESPRIT, cf. Sec. IV B, and EB-DETECT, cf. Sec. IV C, were evaluated in a real moderately reverberated room of size 5.8×5.9 m, exhibiting a reverberation time of about $T_{60}=250$ ms. The array shown in Fig. 10 was placed at the center of a 48-element circular loudspeaker array of 1.5 m radius,[29,30] thereby offering the possibility to simulate 48 source positions in the full field of view of the microphone array.

The performance of EB-ESPRIT due to a *single* speech signal successively applied to all 48 loudspeakers is shown in one superimposed graph in Fig. 12. Here, EB-ESPRIT was evaluated in the time domain, where 100 ms of data was combined into one block. The recorded wavefield, containing a speech signal, was decomposed up to order $\mathcal{N}=3$ and the operating bandwidth was chosen to cover the frequencies from 300 Hz to 3000 Hz. Note that by considering this frequency range, near-field effects can be neglected and plane-wave propagation can be assumed for this particular setup.[17] As can be seen, the maximum error in the estimated azimuth angle, i.e., $\varphi - \hat{\varphi}$, is always less than $3°$ suggesting a highly accurate estimation of the DOA of the impinging plane-wave. Note that no speech activity detector has been employed and that the larger variances around Block Nos. 10

FIG. 12. Measured DOA estimation performance of EB-ESPRIT for $\mathcal{L}=1$.

and 40 are due to speech pauses present in the signal.

The next set of evaluations examines the performance of EB-ESPRIT due to *two simultaneously* impinging plane waves, see Fig. 13. In this case, however, two uncorrelated white-noise signals were chosen since EB-ESPRIT, as with all subspace-based methods, requires the number of active sources in each signal block as *a priori* information. Although this parameter can be obtained via the EB-DETECT algorithm, is kept fixed here in order to avoid possible errors introduced by EB-DETECT. For the purpose of evaluating EB-ESPRIT for two sources, one signal is kept fixed at Loudspeaker No. "1", while the other one is successively fed to one of the other 47 loudspeakers. Additional cylindrically isotropic noise[24] of varying levels is applied in order to test the algorithm's robustness against noise. Figure 13(a) reproduces the mean estimation error in degrees while Fig. 13(b) shows the estimator variance in dB of EB-ESPRIT applied to the source played back by Loudspeaker No. "1". As can be seen, low mean estimation error combined with low estimation variance can be obtained a signal-to-noise ratio (SNR) greater than 0 dB, where "sufficient" spatial distance between the two active loudspeakers is maintained. It is apparent that if the two signals are closely spaced or if the SNR is too low, the performance deteriorates. The counterintuitive effect that for very closely spaced sources the performance seems to improve can be explained as follows. If the two sources are too closely spaced, the algorithm considers them to be a single source. The estimated source position lies right in between the true source positions, with rather remarkable accuracy. The position estimates pertaining to the second
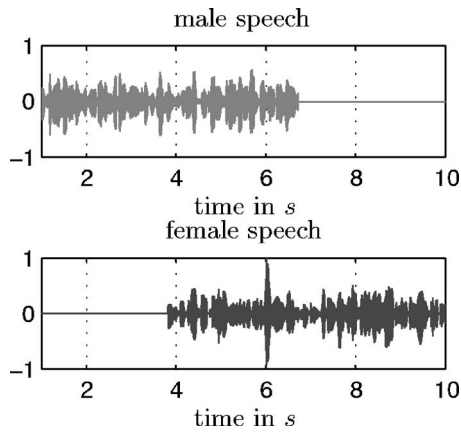
FIG. 14. System excitation signals for EB-DETECT.

(a) $\mathcal{L} = 2$ (b) $\mathcal{L} = 3$

(c) $\mathcal{L} = 4$ (d) $\mathcal{L} = 5$

FIG. 16. Snapshots of the real-time DOA estimation implementation using EB-ESPRIT for $\mathcal{L} > 1$.

source exhibit comparable performance and are therefore not reproduced here.

The final set of performance evaluations pertains to the EB-DETECT algorithm. As detailed earlier, the goal here is to estimate the number of currently active signals in a wavefield as an input parameter for EB-ESPRIT. Here, the two excitation signals are the speech signals shown in Fig. 14. The male speech signal is reproduced by Loudspeaker No. "5", i.e., $\varphi_1 \approx \pi/6$, while the female speech signal is played back by Loudspeaker No. "17", i.e., $\varphi_2 \approx 2\pi/3$. The signals are recorded by the baffled circular microphone array shown in Fig. 10. The data samples are combined into data blocks of 250 ms length. The maximum wavefield decomposition order is $\mathcal{N} = 5$ and $(kR)_{max} \approx 5.2$. The performance of EB-DETECT is shown in Fig. 15. This figure, in essence, presents a birds-eye view of the adaptively estimated modal impulse response of the plane wave, cf. "pw" Fig. 9, as it progresses in time. The two parallel horizontal white lines bound the region of the modal impulse response that has to be taken into account for the detection purpose, cf. Fig. 9. In combination with Fig. 14, peaks corresponding to regions of a single active speech signal as well as regions of simultaneously active speech signals can be identified, thus allowing for an estimation of the number of active sources in a wavefield. Spurious peaks not corresponding to either source can be removed by appropriate postprocessing due to their small relative amplitudes.

FIG. 15. Performance of EB-DETECT, where $\varphi = [\pi/6, 2\pi/3]^T$.

VI. CONCLUSIONS

In this paper, a decomposition of wavefields by circular apertures has been described. The obtained components have been coined as EBs. It has been shown that a relatively short acoustically rigid cylinder can be approximated reasonably well by an infinitely long rigid cylindrical scatterer with satisfactory accuracy. One of the central messages of this paper is that traditional subspace-based direction finding algorithms, which suffer from narrowband assumptions regarding the source signals, can be directly derived in the modal domain, thereby allowing for the localization of multiple simultaneously active acoustic wideband signals. Also, a method based on modal synthesis has been described for estimating the number of active sources in a wavefield. All methods have been validated by measurements made with a compact-sized realization of a circular microphone array mounted into an acoustically rigid cylindrical baffle. It is important to note that all algorithms presented here are real-time capable. In fact, a real-time capable EB-ESPRIT system has been realized. Figure 16 depicts screenshots of a real-time demonstrator operating on up to five simultaneously active wideband noise sources using the 48-element loudspeaker array. Note that the derivation of all algorithms presented in this paper are based on the assumption of free-field propagation. Although satisfactory performance has been shown for moderately reverberated environments, future developments—which should explicitly model reverberation—promise improved performance for highly reverberated environments.

ACKNOWLEDGMENT

[1]H. L. V. Trees, *Optimum Array Processing, Part IV of Detection, Estimation, and Modulation Theory* (Wiley-Interscience, New York, 2002).

[2]D. Johnson and D. Dudgeon, *Array Signal Processing—Concepts and Techniques* (Prentice-Hall, Englewood Cliffs, New Jersey, 1993).

[3]M. Brandstein and D. Ward, eds., *Microphone Arrays—Signal Processing Techniques and Applications* (Springer, Berlin, Heidelberg, New York, 2001).

[4]V. MacDonald and P. Schultheiss, "Optimum passive bearing estimation in a spatially incoherent noise environment," J. Acoust. Soc. Am. **46**, 37–45 (1969).

[5]C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **24**, 320–327 (1976).

[6]J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," J. Acoust. Soc. Am. **107**, 384–391 (2000).

[7]H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann, "Simultaneous localization of multiple sources using blind adaptive mimo filtering," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing*, III-97–III-100, Philadelphia, PA, (2005).

[8]H. Teutsch and W. Kellermann, "EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing*, III-89–III-92 (2005).

[9]H. Teutsch and W. Kellermann, "Estimation of the number of acoustic wideband sources using eigenbeam processing for circular apertures," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 110–113 (2005).

[10]H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," *Proc. Seventh International Symposium on Signal Processing and its Applications (ISSPA 2003)*, Vol. **2**, 411–414 (2003).

[11]E. Claudio and R. Parisi, "Multisource localization strategies," *Microphone Arrays—Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward, Chap. 9, 181–201 (Springer, Berlin, Heidelberg, New York, 2001).

[12]I. Potamitis, H. Chen, and G. Tremoulis, "Tracking of multiple moving speakers with multiple microphone arrays," IEEE Trans. Speech Audio Process. **12**, 520–529 (2004).

[13]R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," IEEE Trans. Speech Audio Process. **37**, 984–995 (1989).

[14]J. Meyer and G. Elko, "A highly scalable spherical microphone array based on a orthonormal decomposition of the soundfield," *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, II-178–II-1784 (2002).

[15]P. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).

[16]E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography* (Academic, London, 1999).

[17]H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition* (Springer, Berlin, Heidelberg, New York, 2007).

[18]M. Junger and D. Feit, *Sound, Structures, and Their Interaction* 2nd Ed. (MIT Press, Cambridge, MA, 1986).

[19]E. Weisstein, *CRC Concise Encyclopedia of Mathematics*, 2nd Ed. (Chapman and Hall/CRC, Boca Raton, FL, 2003).

[20]M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions* (Dover New York, 1972).

[21]A. Oppenheim and R. Schafer, *Discrete- Time Signal Processing*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1998).

[22]C. Mathews and M. Zoltowski, "Eigenstructure techniques for 2D angle estimation with uniform circular arrays," IEEE Trans. Signal Process. **42**, 2395–2407 (1994).

[23]Z. Li and R. Duraiswami, "A robust and self-reconfigurable design of spherical microphone array for multi-resolution beamforming," *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, IV-1137–IV-1140 (2005).

[24]G. Elko, "Differential microphone arrays," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, edited by Y. Huang and J. Benesty, Chap. 2, 11–65 (Kluwer, Dordrecht, 2004).

[25]G. H. Golub and C. F. V. Loan, *Matrix Computations*, 2nd Ed. (Johns Hopkins University Press, Baltimore, MD, 1989).

[26]M. Haardt and J. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," IEEE Trans. Signal Process. **43**, 1232–1242 (1995).

[27]R. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **34**, 276–280 (1986).

[28]S. Haykin, *Adaptive Filter Theory*, 4th Ed. (Prentice Hall, Englewood Cliffs, NJ, 2002).

[29]S. Spors, H. Buchner, and R. Rabenstein, "Efficient active listening room compensation for wave field synthesis," *Proc. 116th Audio Engineering Society (AES) Convention*, Berlin, Germany (2004).

[30]"Chair of multimedia communications and signal processing, University of Erlangen-Nuremberg, Germany," ⟨http://www.LNT.de/lms⟩.

Adaptive focusing for transcranial ultrasound imaging using dual arrays

F. Vignon, J. F. Aubry,[a] M. Tanter, A. Margoum, and M. Fink

Laboratoire Ondes et Acoustique, ESPCI, Université Paris VII, C.N.R.S U.M.R.7587,
10 rue Vauquelin, 75231 Paris Cedex 05, France

Ultrasonic brain imaging remains difficult and limited because of the strong aberrating effects of the skull (absorption, diffusion and refraction of ultrasounds): high resolution transcranial imaging would require adaptive focusing techniques in order to correct the defocusing effect of the skull. In this paper, a noninvasive brain imaging device is presented. It is made of two identical linear arrays of 128 transducers located on each side of the skull. It is possible to separate the respective influence of the two bone windows on the path of an ultrasonic wave propagating from one array to the other, and thus estimate at each frequency the attenuation and phase shift locally induced by each bone window. The information obtained on attenuation and phase is used to correct the wave fronts that have to be sent through the skull in order to obtain a good focusing inside the skull. Compared to uncorrected wave fronts, the spatial shift of the focal spot is corrected, the width of the focal spot is reduced, and the sidelobes level is decreased up to 17 dB. Transcranial images of a phantom are presented and exhibit the improvement in image quality provided by this new noninvasive adaptive focusing method. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354073]

I. INTRODUCTION

Transcranial ultrasonic brain imaging on adults is currently limited by the strong aberrating effect of the skull bone. In order to obtain a good ultrasound image, it is necessary to focus ultrasound beams in the area under investigation: the thinner the focal spot, the better the image resolution; the lower the sidelobes[1] level, the better the contrast. White *et al.*[1] first investigated the defocusing effect of the skull that prevents one from obtaining good ultrasound images of the brain: the skull bone is absorbing the ultrasonic wave and the speed of sound is not homogeneous inside the skull, so that the phase of the wave front is locally shifted.

As the speed of sound and absorption inside the skull are spatially dependent[1,2] and are *a priori* unknown, it is necessary to use adaptive methods to focus through the skull. Phillips *et al.*[3] and Miller Jones[4] first suggested to apply a time-shift compensation in order to correct the phase distortions induced by the skull. However, this approach is limited by a major drawback: in the limited frequency range available for transcranial brain imaging (1.5–3 MHz), the thickness of the skull bone induce a strong aberration that cannot be reduced to a simple time-delay law. In such strongly aberrating configurations, Fink *et al.*[5] demonstrated that time reversal of ultrasonic waves is much more effective. However, it does not totally correct the skull aberrations because of ultrasonic absorption in the bone that breaks time reversal symmetry of the wave equation.[6] A first improvement of this technique consisted in applying an amplitude compensation of time reversed signals before emission.[6,7] This gain compensation acts as an optimal correction technique for absorption effects as long as dissipation remains spatially located in a very thin absorbing layer. Although this first refinement improves the focusing quality, this assumption of thin absorbing layer is not valid in many areas of the skull bone. In the continuity of the works done by Ebbini *et al.*,[8–11] a more general approach proposed by Tanter *et al.*[12] allowed us to reach the optimal focusing by building a complete inverse filter of the transcranial propagation both in space and time domains. This spatiotemporal inverse filter (the STIF method) can be computed from the experimental measurement of the impulse responses relying on the array to a set of receivers in the focal area. The STIF technique was applied to transcranial focusing by Aubry *et al.*, who demonstrated that a perfect correction of skull aberrations could be achieved.[13] Recently, a faster and more robust extension of this technique based on the concept of iterative time reversal between two arrays was introduced by Montaldo *et al.*[14] However, these methods based on the STIF concept are totally unrealistic in practical configurations as they require the presence of a second array of transducers in the focal area (inside the brain!) and thus cannot be used for medical imaging.

In the current study, a new adaptive strategy overcoming this major drawback is proposed. The second array is no longer located inside the brain but is moved outside the head. A similar setup was proposed two decades ago by Phillips *et al.*[3] and Miller Jones[4] for noninvasive time-shift correction: they proposed to use one single transducer on one side of the head emitting impulses that were recorded and processed by an array located on the other side of the head. This paper can thus be seen as an extension of the work done by Philips *et al.* in the 1980s concerning the noninvasive approach and the work done by Tanter *et al.* in the past ten years concerning the aberration correction.

---
[a]Electronic mail: jean-francois.aubry@loa.espci.fr
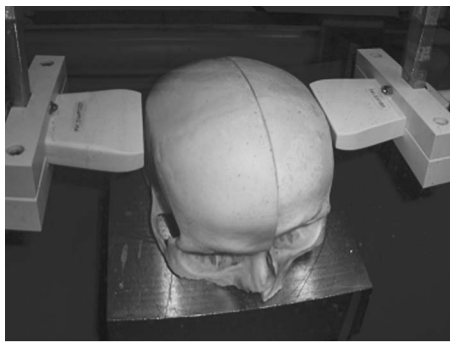
(a)
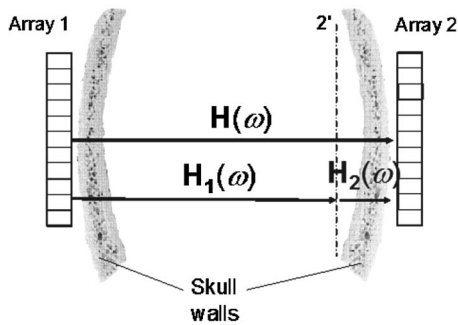
(b)

FIG. 1. (a) Picture of the experimental setup. Two half skulls can be placed alternatively or together between two arrays of transducers. (b) Schematic representation of the experimental set, corresponding to (a), along with an illustration of the physical meaning of the matrices $\mathbf{H}_1(\omega)$ (used in Secs. III and IV), $\mathbf{H}(\omega)$, and $\mathbf{H}_2(\omega)$ (used in Sec. IV).

Thanks to the use of these dual arrays surrounding the patient head (Fig. 1), a noninvasive extension of the STIF technique is proposed and tested. This adaptive focusing procedure is achieved in two steps: the STIF technique is first applied between both arrays and permits one to focus optimally through the whole head from one array to each transducer of the second one. In a second step, the aberrations induced by the skull area located close to the second array are estimated and used to migrate the focal spots inside the brain. In the following, this new strategy will be referred to as the Migrated STIF technique.

The paper is organized as follows: in Sec. II, the defocusing effect of the skull is quantitatively evaluated. In Sec. III, the STIF technique is briefly reintroduced and its potential for spatial focusing through the skull is evaluated. The noninvasive extension of the STIF method is presented in Sec. IV. The experimental validation of this Migrated STIF method (M-STIF) is shown in Sec. V. A single focusing experiment is first analyzed, then the M-STIF method is validated for spatial focusing through different specimens of human skulls. Finally, image quality enhancement is demonstrated on *ex vivo* transcranial images of phantoms.

II. DEFOCUSING EFFECT OF THE SKULL BONE

The experimental setup presented in Fig. 1 is used throughout this study: a sagitally cut human skull is located between two identical linear arrays of 128 ultrasonic trans-
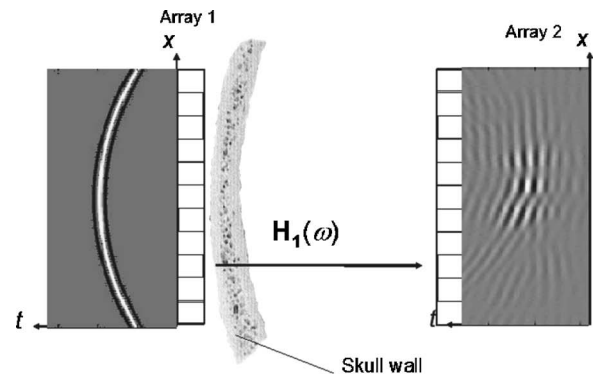
(a)
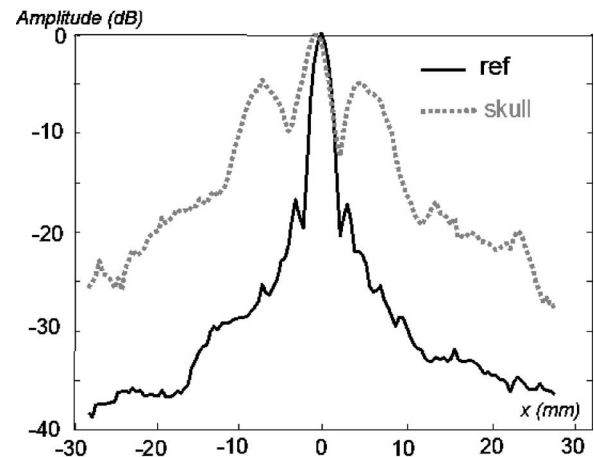
(b)

FIG. 2. Evaluation of the defocusing effect of the skull. (a) A cylindrical wave front and the resulting spatiotemporal focusing through a skull wall. Bright zones correspond to high pressures, dark zones to low ones. (b) Lateral scans of the temporal maxima of the pressure in focal plane, in dB. Black curve: focusing a cylindrical wave front through water (reference experiment), gray dotted curve: focusing through the skull.

ducers. The whole setup is immersed in water. All skull samples used in this study have been cleaned from organic matter and degassed.

The central frequency of the transducers is 1.5 MHz, with a 50% bandwidth. Each element is connected to its own Input/Output fully programmable electronic channel. The array pitch is 0.5 mm so that the lateral aperture of each array is 6.4 cm. The distance between the arrays is of the order of 15 cm. The elevation dimension of the transducers is 1.2 cm (twelve times the mean wavelength in water).

In this section only one half of the skull is placed between the arrays, as illustrated in Fig. 2(a). The defocusing effect of the skull is experimentally evaluated by emitting a cylindrical law with array 1 in order to focus on the central element of array 2. This cylindrical law and the spatiotemporal acoustic field consequently recorded by array 2 in the focal plane are presented in Fig. 2(a). In order to get a quantitative idea of the spatial focusing quality, the temporal maximum of the acoustic pressure received in focal plane is scanned along the x direction of Fig. 2(a) and its spatial variations are plotted in decibels in Fig. 2(b) (gray dotted curve). In Fig. 2(b) is also plotted as a reference the directivity pattern of the focal spot obtained by emitting a cylin-

drical wave front from array 1 to array 2 in a homogeneous medium (black curve). When focusing the cylindrical law through the skull, the maximum pressure amplitude is laterally shifted 1 mm to the left of the targeted point, the main lobe is two times wider (5 vs 2.5 mm) and the sidelobe level is more than 20 dB higher than in the homogeneous medium. Such alterations of the focusing pattern will respectively induce distortion, resolution, and contrast losses on transcranial ultrasound images.

III. THE (INVASIVE) SPATIOTEMPORAL INVERSE FILTER

A complete description of the STIF theory has been done by Tanter et al.,[12] and detailed experiments of invasive transcranial focusing with the STIF method can be found in Ref. 13. In the following, the invasive STIF technique and formalism are briefly reintroduced, in order to better understand the noninvasive implementation of the Migrated STIF technique.

A. The propagation operator $H_1(\omega)$

Wave propagation from the $N=128$ elements of the imaging array (array 1) to a set of N control points embedded in the medium to image can be described in a matrix formalism. The control points are materialized by a second array of transducers placed in front of array 1 in focal plane, as array 2 in Fig. 2(a). For each pair (control transducer j of array 2, emitting transducer i of array 1), the Green's function $h_{ji}(t)$ is experimentally acquired. It corresponds to the signal received on the jth transducer of array 2 when a temporal Dirac function is emitted from the ith transducer of array 1. This signal includes all the propagation effects through the considered medium as well as the acousto-electric responses of both transducers. Now let $e_i(t)$, $1 \leq i \leq N$ be the input signal on the ith transducer. As the effects of the propagation in the medium are supposed linear, the output signal $f_j(t)$, $1 \leq j \leq N$, received on the jth control point is

$$f_j(t) = \sum_{i=1}^{N} h_{ij}(t) * e_i(t),\tag{1}$$

where the asterisk (*) represents the temporal convolution operator.

This relation is written more easily in the Fourier domain where convolutions become multiplications. Fourier transforming (1) leads to

$$f_j(\omega) = \sum_{i=1}^{N} h_{ij}(\omega) e_i(\omega).\tag{2}$$

N equations similar to Eq. (2) can be written for the N control points and summed up into a single matrix relation:

$$F(\omega) = \mathbf{H}(\omega) E(\omega),\tag{3}$$

where $E(\omega) = {}^t[e_1(\omega), e_2(\omega), \ldots, e_N(\omega)]$ is the emission vector that contains in position i the Fourier transform of the signal emitted from transducer i of array 1; $F(\omega) = {}^t[f_1(\omega), f_2(\omega), \ldots, f_N(\omega)]$ is the focused vector that contains in position j the Fourier transform of the signal re-

ceived on control transducer j of array 2. The transfer matrix $\mathbf{H}(\omega) = \{h_{ji}(\omega)\}_{1 \leq i \leq N, 1 \leq j \leq N}$ describes the propagation through the skull from array 1 to array2 for a chosen frequency component and is called the propagator.

B. The spatiotemporal inverse filter (STIF)

By inverting the matrix relation (3) at each frequency within the bandwidth of the transducers, one can compute the Fourier components $E(\omega)$ of the signals to emit from array 1 in order to focus on any desired target vector $S(\omega)$ on array 2. For instance, in order to focus on a temporal Dirac on the j_0^{th} control point, $S(\omega)$ is set as

$$S_{j0}(\omega) = 1, S_{j \neq j0}(\omega) = 0.\tag{4}$$

The emission signals are then given by

$$E(\omega) = \hat{\mathbf{H}}(\omega)^{-1} S(\omega),\tag{5}$$

where $\hat{\mathbf{H}}(\omega)^{-1}$ is the pseudo-inverse of $\mathbf{H}(\omega)$, regularized by singular value decomposition.[12] The inversion is performed at each frequency within the bandwidth of the transducers. The temporal signals $\{e_i(t)\}_{1 \leq i \leq N}$ to be emitted from array 1 to focus on S on array 2 are finally calculated by an inverse Fourier transform of $E(\omega)$.

Emission signals calculated by STIF in order to adaptively focus through a skull wall are shown in Fig. 3(a) along with the corresponding acoustic field recorded on the receiver array in focal plane. The spatiotemporal emission signals are complicated and nonintuitive but they are able to interfere constructively at the desired focal spot location and destructively in any other position, restoring a focusing quality as good as the one obtained by focusing a cylindrical law through water as can be seen in the lateral focusing patterns presented in Fig. 3(b). The STIF method thus proves that there exists a set of emission signals able to focus through the skull with a focusing pattern close to the one obtained in a homogeneous medium. However it requires the presence of acoustic transducers in the focal region for the acquisition of the propagator: this technique is invasive and is not applicable as is for in vivo imaging.

The noninvasive adaptation of the STIF focusing technique for transcranial imaging is presented in the following.

IV. THE MIGRATED SPATIOTEMPORAL INVERSE FILTER

A. Experimental setup and introduction of the formalism

The experimental setup has to be noninvasive: as shown in Fig. 1 the two arrays (array 1: the imaging array and array 2: the control array) are now located on each side of the full skull. Array 1 will be used later to image the right half of the brain. By switching the role of arrays 1 and 2, array 2 could be used to image the other half of the brain.

For a mathematical description of the method three propagators must be introduced. First, the global propagator $\mathbf{H}(\omega)$ through the entire skull from array 1 to array 2, which can be acquired experimentally noninvasively. Second, $\mathbf{H}_1(\omega)$, which corresponds to the propagator from array 1 to

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Vignon *et al.*: Adaptive focusing for transcranial imaging   2739
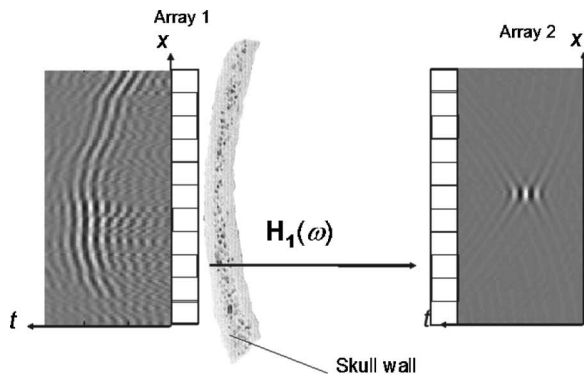
(a)
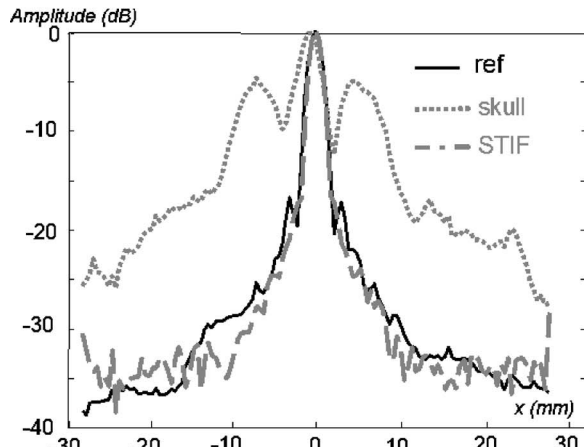
(b)

FIG. 3. Focusing through the skull by STIF. (a) The adaptive focusing wave front calculated by STIF, and the resulting spatiotemporal focusing through the skull. (b) Lateral scan of the temporal maxima of the pressure in focal plane, in dB. Black curve: focusing a cylindrical wave front through water (reference); gray dotted curve: focusing a cylindrical wave front through the skull, and gray dashed curve: focusing the STIF solution through the skull.

a fictive array 2′ inside the skull stuck to the skull bone window in front of array 2, and then $\mathbf{H}_2(\omega)$, which corresponds to the propagator from the fictive array 2′ to array 2. The whole formalism is illustrated in Fig. 1(b).
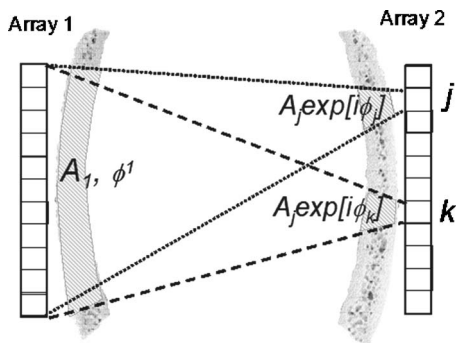
FIG. 4. Schematization of the principle used to estimate the phase and amplitude factors. When a beam propagates from one transducer j or k of array 2 to array 1 or in the reverse sense, it crosses only a small portion of the second skull wall that depends on the emitting transducer. As the beam gets larger during its propagation, it always crosses the same portion of the first skull wall that lies in front of the full aperture of array 1, and the phase factor and amplitude shift that it implies thus do not depend on the emitting transducer position of array 2.
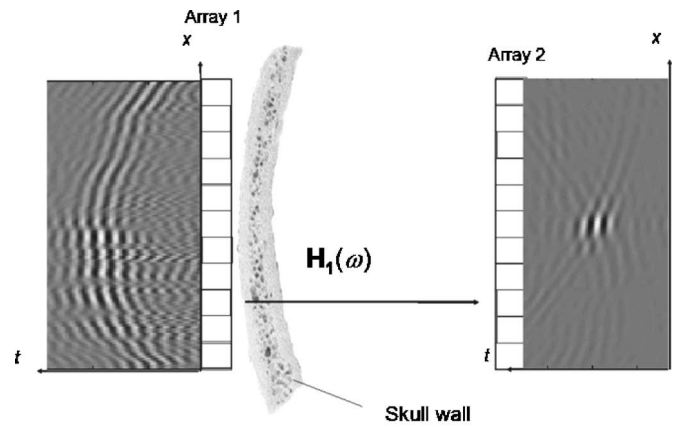
(a)
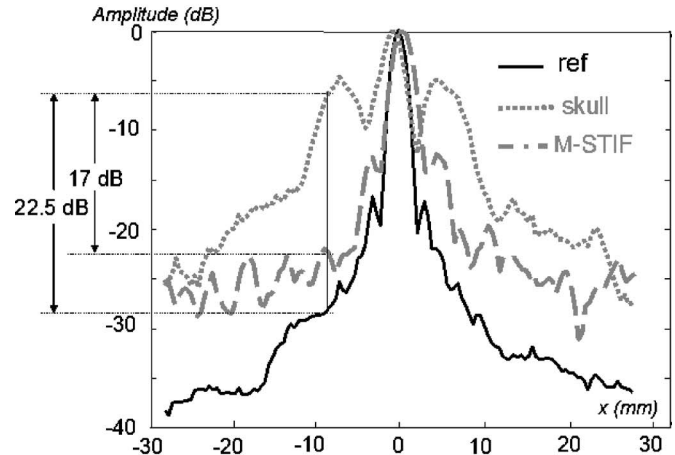
(b)

FIG. 5. Focusing through the skull by the noninvasive migrated STIF. (a) The adaptive focusing wave front calculated by migrated STIF, and the resulting spatiotemporal focusing through the skull. (b) Lateral scan of the temporal maxima of the pressure in focal plane, in dB. Black curve: focusing a cylindrical wave front through water (reference); gray dotted curve: focusing a cylindrical wave front through the skull, and gray dashed curve: focusing the migrated STIF signals through the skull.

The aim now is to deduce from $\mathbf{H}(\omega)$ (the only measurable propagator) an estimation of $\mathbf{H}_1(\omega)$ that could be inverted in order to focus ultrasounds on the plane of the fictive array 2′ virtually inside the brain.

$\mathbf{H}(\omega)$, $\mathbf{H}_1(\omega)$, and $\mathbf{H}_2(\omega)$ are linked by

$$\mathbf{H}(\omega) = \mathbf{H}_2(\omega)\mathbf{H}_1(\omega). \qquad (6)$$

This relation expresses the fact that in order for a wave to propagate from array 1 to 2 [multiplication by $\mathbf{H}(\omega)$] it first has to propagate from array 1 to 2′[multiplication by $\mathbf{H}_1(\omega)$] and then from 2′ to 2 [multiplication by $\mathbf{H}_2(\omega)$].

Instead of trying to directly estimate $\mathbf{H}_1(\omega)$ from information contained in the global propagator $\mathbf{H}(\omega)$, it is proposed here to first estimate $\mathbf{H}_2(\omega)$ by a diagonal propagator $\widetilde{\mathbf{H}}_2(\omega)$ and then invert it in order to deduce an estimator $\widetilde{\mathbf{H}}_1(\omega)$ of $\mathbf{H}_1(\omega)$:
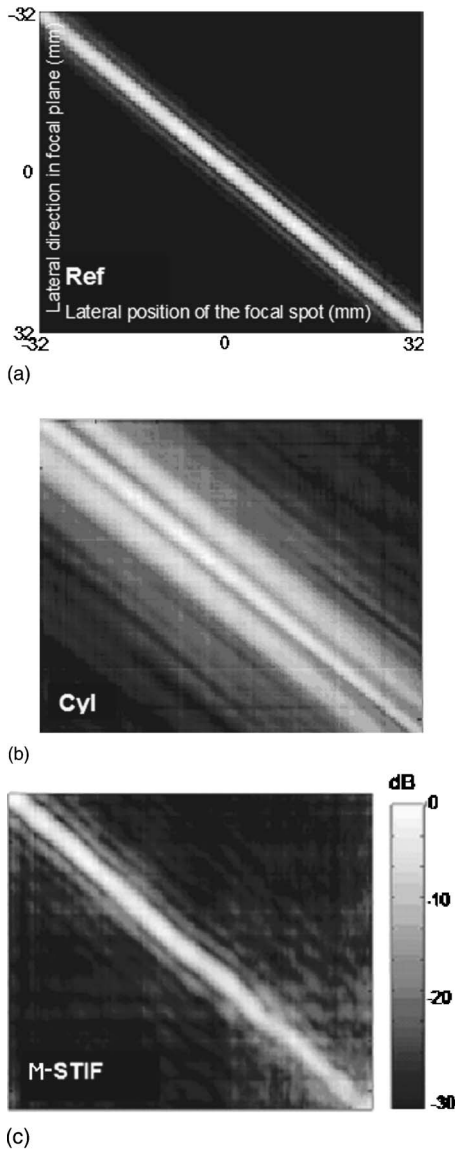
FIG. 6. (a) Lateral scans of the focal spots obtained at each of the 128 focal points by focusing the cylindrical wave fronts through water, in dB. The lateral pressure scan obtained in focal plane when focusing on the jth transducer of array 2 corresponds to the jth column of the represented matrix, the bright zones corresponding to high pressures and the dark zones to low pressures. (b) The lateral scans of the focal spots focusing cylindrical wave fronts through the first skull wall. (c) Focusing of the migrated STIF signals through the first skull wall.

$$\widetilde{\mathbf{H}}_1(\omega)\widetilde{\mathbf{H}}_2(\omega)^{-1}\mathbf{H}(\omega). \tag{7}$$

This indirect method allows us to keep in the estimator $\widetilde{\mathbf{H}}_1(\omega)$ the information on the near field of array 1 that is contained in $\mathbf{H}(\omega)$.

Reducing $\mathbf{H}_2(\omega)$ to a diagonal propagator physically corresponds to the assumption that the skull wall is infinitely thin, and located close to array 2. $\widetilde{\mathbf{H}}_2(\omega)$ is then written as a diagonal propagator in which element (j,j) is an amplitude and phase factor describing the acoustical effects of the small portion of skull bone in front of the jth element of array 2:

$$\widetilde{h}_{2ij}(\omega) = 0 \quad for \ i \neq j,$$

$$\widetilde{h}_{2jj}(\omega) = A_j(\omega)\exp(i\phi_j(\omega)). \tag{8}$$

$A_j(\omega)$ and $\phi_j(\omega)$ are the amplitude and phase factors by which is affected an ultrasound beam that crosses the portion of skull wall in front of transducer j of array 2. Those factors are obtained by the analysis of the propagator through the entire skull $\mathbf{H}(\omega)$ as explained in the following section.

B. Estimation of the amplitude and phase factors induced by the second skull wall

1. Estimation of the amplitude factor

Concerning the amplitude, the influence of the portion of skull located in front of transducer j of array 2 can be modeled by an amplitude factor $A_j(\omega)$ which multiplies the Fourier component ω of the field crossing this local area of the skull. This amplitude factor is estimated as follows. A signal of normalized amplitude 1 is first emitted by transducer j of array 2. The transmitted wave is received on array 1, time-reversed, and sent back by all elements of array 1 to array 2. The amplitude $A_j^{TR}(\omega)$ of the Fourier component at frequency ω of the signal then received on receiver j of array 2 is proportional to the square of $A_j(\omega)$ as the wave has crossed the skull bone twice during the round trip:

$$A_j^{TR}(\omega) = A^1(\omega)^2 A_j(\omega)^2. \tag{9}$$

The proportionality factor $A^1(\omega)^2$ characterizes the global absorption induced by the whole first skull bone window in front of array 1 (it is squared because the wave crosses the first half skull two times during the round trip) and is assumed not to depend, or very slightly, on j. Indeed, when shooting from or focusing on different transducers j and k of array 2, the beam crosses two different small portions of the skull wall in front of array 2 and thus is affected by two different amplitude factors $A_j(\omega)$ and $A_k(\omega)$, while it is af

TABLE I. −6 dB width of the focal spot with the cylindrical law through the skull, M-STIF technique through the skull, and reference focusing (cylindrical law through a homogeneous medium) for all eight experiments. Also indicated is its mean value and standard deviation over the eight experiments.

	Exp. 1	Exp. 2	Exp.	Exp. 4	Exp. 5	Exp. 6	Exp. 7	Exp. 8	Mean	S.D.
Cylindrical	5	4.5	4	3	5.5	9	3.5	4.5	**4.9**	1.8
M-STIF	3	3.5	3	3	3.5	3.5	3.5	4	**3.4**	0.4
Reference	2.5	3	2.5	2.5	2.5	2.5	2.5	3	**2.6**	0.2

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Vignon *et al.*: Adaptive focusing for transcranial imaging   2741

FIG. 7. Estimation of the focal spot parameters as a function of the lateral position in focal plane, when focusing a cylindrical wave front through water (reference experiment, black curves), focusing a cylindrical law through the skull (gray dotted curves), focusing the M-STIF signals through the skull (gray dashed curves). (a) –6 dB width of the focal spot; (b) lateral shift of the focal spot, (c) contrast loss with respect to the reference.

fected by the whole surface of the skull in front of array 1 experiencing nearly the same amplitude factor $A^1(\omega)$, as evidenced by the dotted and dashed lines in Fig. 4.

As only the relative amplitude factors induced by different parts of the second skull wall are needed, the proportionality factor $A^1(\omega)^2$ due to the crossing of the first skull wall is of no importance. The estimator $\widetilde{A}_j(\omega)$ of the amplitude $A_j(\omega)$ corresponding to the portion of second skull wall located in front of transducer j of array 2 can be taken as the square root of the amplitude of the received signals after this time-reversal process:

$$\widetilde{A}_j(\omega) = [A_j^{TR}(\omega)]^{1/2}. \tag{10}$$

2. Estimation of the phase factor

The phase shift $\phi_j(\omega)$ affecting the Fourier component ω of a wave crossing the portion of skull in front of transducer j of array 2 is estimated as follows. The cylindrical wave front that would converge on transducer j of array 2 in the absence of the skull is sent from array 1. The phase $\psi_j(\omega)$ of the signal received on transducer j after crossing the full skull is extracted at each frequency. This phase differs from $\phi_j(\omega)$ by an additional phase factor $\phi^1(\omega)$ that comes from the crossing of the first half of the skull:

$$\psi_j(\omega) = \phi^1(\omega) + \phi_j(\omega). \tag{11}$$

However the additional phase factor $\phi^1(\omega)$ does not, or very slightly, depend on the position j of the second array we are focusing on. Indeed, when focusing on two different transducers j and k of the second array the beam crosses two different small portions of the skull wall in front of array 2 and thus is affected by two different phase shifts $\phi_j(\omega)$ and $\phi_k(\omega)$, while it is affected by the whole surface of the skull in front of array 1 experiencing nearly the same phase shift $\phi^1(\omega)$, as evidenced by the dotted and dashed lines in Fig. 4. As only the relative phase shifts induced by different parts of the second skull are needed, the additional constant $\phi^1(\omega)$ is of no importance and a phase shift estimator $\widetilde{\phi}_j(\omega)$ for the phase shift $\phi_j(\omega)$ can be taken as

$$\widetilde{\phi}_j(\omega) = \psi_j(\omega). \tag{12}$$

C. Noninvasive focusing

Once $A_j(\omega)$ and $\phi_j(\omega)$ $(1 \leq j \leq 128)$ are estimated, $\widetilde{\mathbf{H}}_2(\omega)$ is constructed as indicated in Eq. (8), replacing $A_j(\omega)$ and $\phi_j(\omega)$ by their estimators $\widetilde{A}_j(\omega)$ and $\widetilde{\phi}_j(\omega)$. The estimated propagator $\widetilde{\mathbf{H}}_1(\omega)$ approaching $\mathbf{H}_1(\omega)$ is then given by inversion of $\widetilde{\mathbf{H}}_2(\omega)$ as indicated in Eq. (7).

$\widetilde{\mathbf{H}}_1(\omega)$ has been derived totally noninvasively through the analysis of signals that cross the full skull. It can now be pseudo-inversed to calculate the STIF signals to focus through the first skull wall. Those noninvasive migrated STIF signals $E_{M\text{-STIF}}(\omega)$ are thus given by

$$E_{M\text{-STIF}}(\omega) = \widetilde{\hat{\mathbf{H}}}(\omega)^{-1} S(\omega). \tag{13}$$

TABLE II. Lateral shift of the focal spot with the cylindrical law through the skull and M-STIF technique through the skull for all eight experiments. Also indicated is its mean value and standard deviation over the eight experiments.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7	Exp. 8	Mean	S.D.
Cylindrical	1	1.5	1	1.5	3	0.5	0	0	**1.1**	1.0
M-STIF	0.5	0.5	0.5	0	0	0	2	0	**0.4**	0.7

The temporal emission signals are finally computed via an inverse Fourier transform.

Finally, from the knowledge of this set of emission vectors focusing in the brain close to the inner part of the skull close to array 2, other focal depths can be reached by simply assuming that the speed of sound is constant inside the brain: simple delays can be applied to the original M-STIF signals that can be used to steer the beams deeper inside the brain.

V. RESULTS

A. Analysis of a single experiment

Concretely, the experimental sequence is the following: the propagator $H(\omega)$ is first acquired through the entire skull, then the migrated STIF (M-STIF) emission signals are calculated as described in the previous section. Then in order to test quality of the M-STIF technique, the focusing pattern is scanned by simply removing the portion of skull bone in front of array 2 without moving the portion of skull bone in front of array 1. The M-STIF signals are emitted from array 1 and the corresponding field in the focal plane is recorded by array 2. The M-STIF spatiotemporal signals set to focus on the central position and the corresponding spatiotemporal field received in focal plane are presented in Fig. 5(a). The lateral directivity pattern of the acoustic amplitude received in the focal plane obtained with the M-STIF technique through the first skull wall is plotted in Fig. 5(b) in a decibel scale (dashed gray curve). It is compared to the reference lateral focusing (focusing a cylindrical wave front through water, black curve) and to the distorted noncorrected focusing (cylindrical wave front through the first skull wall, dotted gray curve). The M-STIF technique corrects the lateral shift of the focal spot induced by the skull bone, narrows the main lobe, and the sidelobe level is reduced up to 17 dB lower than without any correction being performed. This is a significant improvement although the sidelobe level is still 5.5 dB higher than in the reference experiment: the focusing quality that could be obtained through water is not reached because of the imperfect estimation of the propagator $H_1(\omega)$. Assuming the diagonality of the propagator $H_2(\omega)$ through the second skull wall leads to a nonoptimal estimation of $H_1(\omega)$. However, an iterative process could be applied in

order to refine the modeling of $H_2(\omega)$ by introducing a band diagonal matrix instead of a diagonal matrix. This work is currently under investigation.

The migrated STIF technique can be used to focus on any point in the focal plane: being able to focus on different positions in the focal area is essential for imaging. In order to validate this experimentally, the M-STIF technique is now used to focus on any of the 128 transducers of array 2. Each focal spot is scanned in order to check the lateral focusing quality for all focusing locations. The focal spots are presented in Fig. 6: on each of the matrices presented in Fig. 6, the ith column is the lateral focusing pattern when focusing on the ith transducer of array 2 (white corresponds to the highest pressures and black to the lowest pressures). For instance the 64th column of the matrices presented in Fig. 6 corresponds to the pressure curves in Fig. 5(b). The compensation of the spatial shift of the focal spot, the narrowing of the main lobe, and a decrease in the sidelobe level are achieved for each focusing position when using the M-STIF technique. Figures 6(c) and 7 show that the correction reached when using the M-STIF technique to focus through the skull is most efficient when focusing at central positions in the focal plane. In this experiment the focusing quality gets worse when focusing on abscisses $x > 20$ mm. Indeed, because of the curvature of the skull, in this region the transducers of array 2 can be as far as 1 cm away from the second skull wall thus damaging the estimation of $H_2(\omega)$. The estimation of the M-STIF emission signals to focus on these positions is thus less valid.

B. Validation of the M-STIF technique on different skull specimen

The same experiment has been repeated using different skulls, focusing through different parts of the first skull wall, in order to test the robustness of the method.

Eight focusing experiments have been performed through different areas of three different degassed and cleaned skulls. For each experiment, focusing was performed on 128 positions in focal plane. The focusing by migrated STIF through the first skull wall is compared to the one obtained with a prefocused cylindrical wave front through

TABLE III. Contrast loss with respect to the reference focusing when using the cylindrical law through the skull and M-STIF technique through the skull for all eight experiments. Also indicated is its mean value and standard deviation over the eight experiments.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7	Exp. 8	Mean	S.D.
Cylindrical	22.5	11	9	17.5	13	12	15	21	**15.1**	4.8
M-STIF	5.5	3.5	5	6.5	2	5	6	9	**5.3**	2.1

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Vignon *et al.*: Adaptive focusing for transcranial imaging   2743
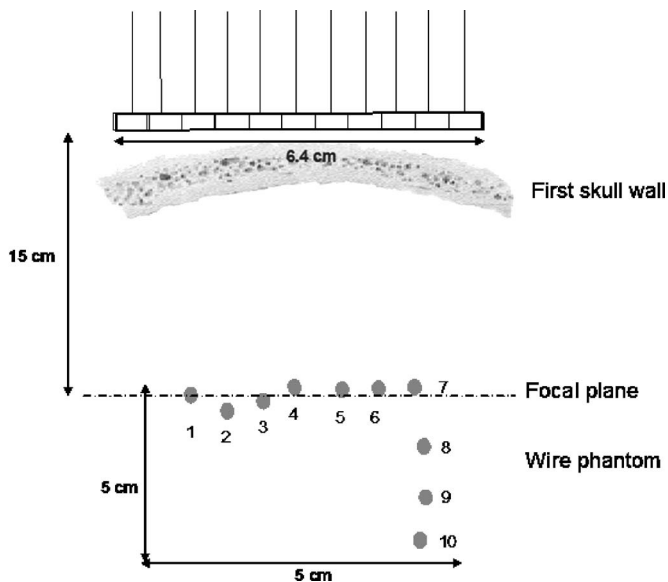
FIG. 8. Schematization of the experimental imaging setup and the wire phantom. The wires have been indexed from 1 to 9. The relevant dimensions of the experiment have been indicated.

water (reference focusing) and the one obtained with a cylindrical wave front through the skull (denoted as cylindrical focusing). The characteristics of the lateral focusing are reported when focusing at the central position in focal plane. Three parameters have been measured:

(1) The −6 dB width of the main lobe of the focal spot.
(2) The lateral shift of the main lobe of the focal spot.
(3) The contrast loss with respect to the reference focusing. For the cylindrical focusing through the skull it is the maximum difference, in decibels, of the sidelobe level between the cylindrical focusing through the skull and the reference focusing. For the M-STIF focusing pattern it corresponds to the difference between the M-STIF sidelobe through the skull and the reference sidelobe at the position of this maximum difference [see Fig. 5(b)].

Experiments 1–3 are performed on a first adult skull, experiments 4–6 are performed on a second adult skull specimen, experiments 7 and 8 are performed on a child's skull (12 years old). On each skull different experiments have been done because different relatively planar areas of the skull adequate for performing the M-STIF technique have been encountered. Those planar areas are always located close to the temple. Because the radius of curvature of the child's skull is smaller it is more difficult to find a planar area on this skull that covers the full aperture of the array. The application of the M-STIF technique is thus most difficult when using the child's skull.

The results are summarized in Table I for the −6 dB width, Table II for the lateral shift, and Table III for the contrast loss, for all eight experiments. In all experiments, the lateral width of the focal spot is smaller with the M-STIF method than with the cylindrical focusing through the skull. In all experiments except the seventh one the M-STIF method corrects partially or totally the lateral shift of the focal spot, and in all experiments the M-STIF method sig-
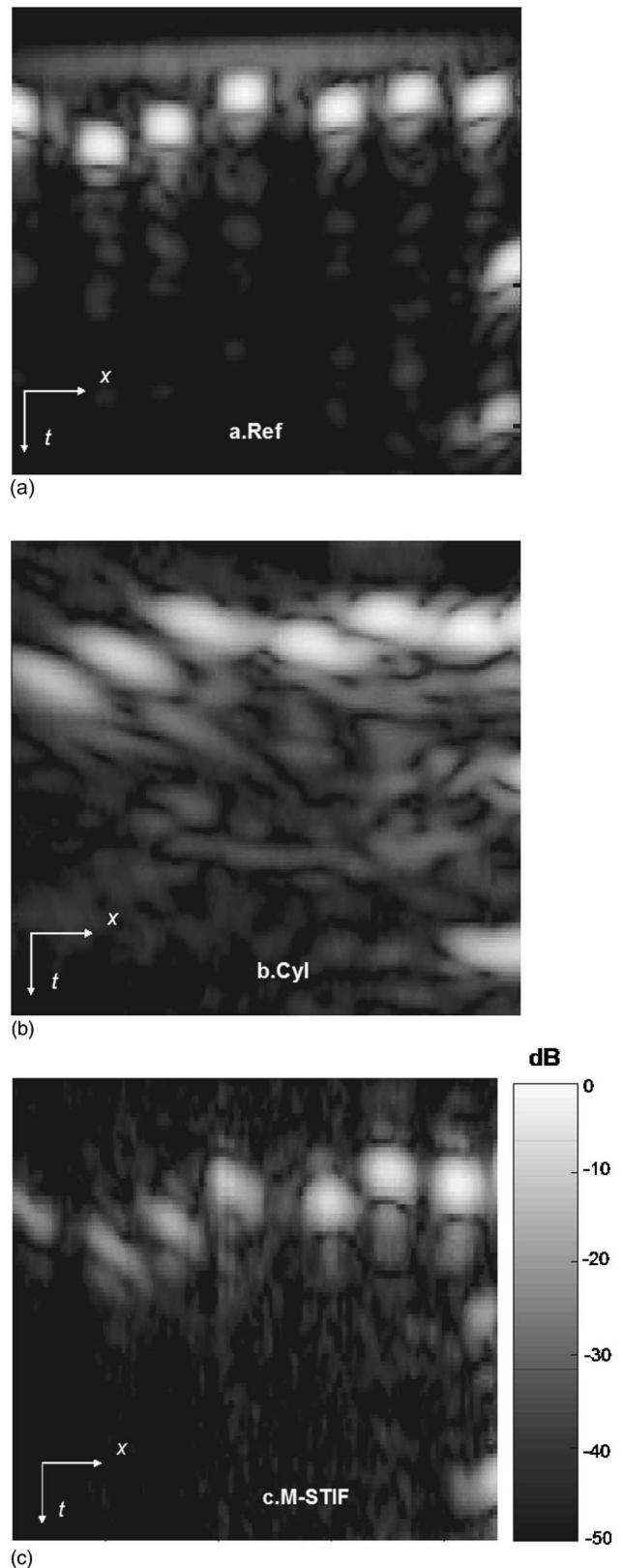
FIG. 9. (a) Reference: image of the wires with cylindrical wave fronts and without the skull (conventional imaging in water). (b) Transcranial image of the wires with the cylindrical wave fronts (conventional focusing). (c) Transcranial image of the wires with the migrated STIF technique.

nificantly reduces the sidelobe level with respect to the side-lobe level of the cylindrical focusing through skull. On average the –6 dB width of the focal spot is lowered by 32%: from 4.9 mm with the cylindrical focusing through the skull to 3.3 mm with the M-STIF through the skull (2.6 mm on average for the reference experiments). The lateral shift of the focal spot is reduced by 63%: from 1.1 to 0.4 mm. Finally, the contrast loss with respect to the reference is lowered by 65%: from 15.1 dB with the cylindrical focusing through the skull to 5.3 dB with the M-STIF method.

In all cases the M-STIF method provides a better lateral focusing quality through the skull than the uncorrected cylindrical wave fronts, and on average the focusing quality reached with the M-STIF method is closer to the reference focusing quality than to the cylindrical focusing quality.

It is worth to say here that experiment 1 corresponds to experiment analyzed in Sec. V A. Tables I–III, show that it is representative of the average results one can obtain applying the M-STIF method on different skulls.

C. Evaluation of the image quality improvement with the migrated STIF method

In order to evaluate the image quality enhancement with the migrated STIF technique, transcranial images of a homemade phantom have been performed *ex vivo*. The phantom is made of a set of nine aluminum wires (0.8 mm diameter) perpendicular to the propagation plane. Those wires are distributed along two axes in order to test axial and lateral focusing quality, and immersed in water (Fig. 8). They have been imaged using cylindrical wave fronts in the absence of the skull [reference image, Fig. 9(a)], using cylindrical wave fronts through the first skull wall [Fig. 9(b)] and using the M-STIF technique through the skull [Fig. 9(c)]. The corrected image exhibits a good enhancement compared to the uncorrected image. In particular, in the uncorrected image [Fig. 9(b)], the shape of the wires is not well rendered, wires 5,6,7 are not resolved and the distortion (the lateral shift of the focal spots) is so strong that wire 1 gets out of the image. The M-STIF image [Fig. 9(c)] resolves wires 5–7, gets wire 1 back into the imaged area, and renders better the shape of the wires. Visually, the general view is much better with M-STIF than with no correction at all.

VI. CONCLUSION

The recent introduction of the STIF for transcranial focusing demonstrated the existence of a set of signals that is able to focus through a human skull with the same quality as conventional prefocused cylindrical signals through a homogeneous medium. The computation of this set of signals required an invasive experimental setup. This technique has been extended in this paper to noninvasive spatiotemporal inverse filter. Thanks to a dual array setup, the optimal set

focusing signals that would be given by the STIF technique can be approached noninvasively. This noninvasive extension of the STIF technique has been termed the Migrated STIF technique because focal spots obtained by STIF outside the head in a first step are migrated inside the brain in a second step. It has been shown that the migrated STIF lowers the -6 dB width of the focal spot by 32%, the lateral shift of the focal spot by 63%, and the contrast loss with respect to the reference by 65% on average. The benefit of such corrections on image quality has been confirmed on transcranial *ex vivo* images of test phantoms.

Applications of precise, adaptive, noninvasive focusing of ultrasounds go well beyond structural imaging. Quantitative blood flow measurements inside the brain can be envisioned with the same approach and could be particularly interesting for the diagnosis of stroke.

[1] D. White, J. Clark, J. Chesebrough, M. White, and J. Campbell, "Effect of skull in degrading the display of echoencephalographic B and C scans," J. Acoust. Soc. Am. **44**, 1339–1345 (1968).

[2] F. Fry and J. Barger, "Acoustical properties of the human skull," J. Acoust. Soc. Am. **65**, 1576–1590 (1978).

[3] D. Phillips, S. Smith, O. Von Ramm, and F. Thurstone, "Sampled aperture techniques applied to B-mode echoencephalography," Acoust. Hologr. **6**, 103–120 (1975).

[4] M. Miller Jones, Ph.D. thesis, Duke University, 1980.

[5] M. Fink, "Time reversed acoustics," Phys. Today **20**, 34–40 (1997).

[6] J.-L. Thomas and M. Fink, "Ultrasonic beam focusing through tissue heterogeneities with a time-reversal mirror; application to transskull therapy," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**(6), 1122–1129 (1996).

[7] M. Tanter, J. L. Thomas, and M. Fink, "Focusing and steering through absorbing and aberrating layers: Application to ultrasonic propagation through the skull," J. Acoust. Soc. Am. **103**(5), 2403–2410 (1998).

[8] E. Ebbini and C. Cain, "Multiple-focus ultrasound phased-array pattern-synthesis: Optimal driving-signal distributions for hyperthermia," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**(5), 540–548 (1989).

[9] R. Seip, P. VanBaren, and E. Ebbini, "Dynamic focusing in ultrasound hyperthermia using implantable hydrophone arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**(5), 706–713 (1994).

[10] J. Shen, H. Erdogan, and E. Ebbini, "An optimal image operator design technique for coded excitation ultrasound imaging system," *Proceedings of IEEE Ultrasonics Symposium*, 1994.

[11] E. Ebbini, L. Pai Chi, and J. Shen, "A new SVD-based optimal inverse filter for ultrasound applications," *Proceedings of IEEE Ultrasonic Symposium*, 1993, pp. 1187–1190.

[12] M. Tanter, J. F. Aubry, J. Gerber, J. L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. I. Basic principles," J. Acoust. Soc. Am. **110**(1), 37–47 (2001).

[13] J. F. Aubry, M. Tanter, J. Gerber, J. L. Thomas, and M. Fink, "Optimal focusing by spatio-temporal inverse filter. II. Application to focusing through absorbing and reverberating media," J. Acoust. Soc. Am. **110**(1), 48–58 (2001).

[14] G. Montaldo, M. Tanter, and M. Fink, "Real time inverse filter focusing through iterative time reversal," J. Acoust. Soc. Am. **115** (2), 220–230 (2004).

Gaussian beams and Legendre polynomials as invariants of the time reversal operator for a large rigid cylinder

Alexandre Aubry,[a] Julien de Rosny,[b] Jean-Gabriel Minonzio,
Claire Prada, and Mathias Fink
Laboratoire Ondes et Acoustique, ESPCI, Université Denis Diderot (Paris VII), CNRS (UMR 7587),
10 rue Vauquelin, 75005 Paris, France

The DORT method (French acronym for decomposition of the time reversal operator) is an active remote sensing technique using an array of antennas for the detection and localization of scatterers. This method is based on the singular value decomposition of the interelement response matrix. In this paper an analytical expression of the singular vectors due to the reflection from a large rigid cylinder is provided. Depending on the array aperture, two asymptotic regimes are described. It is shown that the singular vectors correspond to Hermite-Gaussian modes for large apertures and Legendre polynomials for small ones. Using perturbation theory, the corresponding singular values are deduced. Theoretical predictions are in good agreement with experimental results. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2357700]

I. INTRODUCTION

Acoustical backscattering detection is widely used in many different areas such as medical imaging, nondestructive evaluation, or underwater acoustics. Thanks to technological progresses, arrays made of hundreds of transceivers are built. Consequently, the interest for developing new signal processing techniques is growing fast to make the most of the rich spatial diversity of the collected data. A few years ago an original approach, called the DORT method (French acronym for decomposition of the time reversal operator), was implemented in order to optimize the use of such arrays. The method consists in the singular value decomposition (SVD) of the array response matrix, $\mathbf{K}(\omega)$. Elements of this matrix correspond to the backscattering responses at one frequency ω between all pairs of transducers. The SVD of $\mathbf{K}(\omega)$ provides a set of singular vectors weighted by singular values. The singular vectors of $\mathbf{K}(\omega)$ are the invariants of the time reversal operator $[\mathbf{K}^{\dagger}(\omega)\mathbf{K}(\omega)]$. The SVD is written $\mathbf{K} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger}$, where $\mathbf{\Sigma}$ is a diagonal matrix containing the real positive singular values σ_n. $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices containing the normalized singular vectors $\mathbf{U}_n$ and $\mathbf{V}_n$.

DORT has shown its efficiency in detecting and separating the responses of several scatterers in homogeneous or heterogeneous media.[1] It was applied to the detection of flaws in nondestructive testing,[2] medical acoustics,[3] target detection and identification in underwater acoustics,[4–6] and electromagnetism.[7,8] From the DORT analysis, it is then possible to achieve time reversal imaging thanks to backpropagation algorithms[2,6,9] and also selective focusing.[4,5,8] However, before applying the DORT method to the inverse problem such as target detection and characterization, the direct problem should be investigated. It consists in determining analytically singular vectors and values correspond-

ing to a given target. Until now, almost all such works only address the detection of objects small with respect to the wavelength. For pointlike scatterers, it has been shown that there are only four significant singular values (three for a 2-D configuration) associated to each scatterer[10,11] in acoustics and up to six in electromagnetism.[12]

The case of bigger scatterers has been much less investigated. Chambers[13] extends the analysis of Chambers and Gautesen[10] for a bigger sphere ($ka > 1$, where k is the wave number and a the sphere radius) but singular values and vectors are obtained only numerically. Zhao[14] finds an analytic expression for the first two singular vectors and values in the case of a flat target. His analysis is based on the parabolic approximation of the Green function.

In this study, the DORT method is applied to two different configurations involving a large rigid cylinder. The first configuration corresponds to a "well-resolved" (WR) cylinder, for which the square root product $\sqrt{\lambda a}$ is bigger than the array resolution cell $\lambda F/D$, where λ is the wavelength, F is the distance between the array and the cylinder, and D is the array width. For the second configuration, the cylinder is "poorly resolved" (PR) because its radius a is smaller than the resolution cell ($a < \lambda F/D$). Both models can also be applied to an elastic target if the specular echo dominates the elastic wave radiation such as creeping waves, circumferential waves, etc.

First, experimental singular vectors and values obtained from the two configurations are presented in Sec. II. In Sec. III, these results are compared to those obtained with numerical computations based on a modal decomposition of the scattered wave field. In the next two sections, two asymptotic models explaining the behavior of the singular vectors are proposed. The WR model is deduced from Gaussian beam theory developed in optics. Under the paraxial approximation, the singular vectors are well fitted by Hermite-Gaussian modes. The beam width is established from the ray transfer matrix. In Sec. V, a PR model is proposed and comes from

---

[a]Electronic mail: alexandre.aubry@espci.fr
[b]Electronic mail: julien.derosny@espci.fr
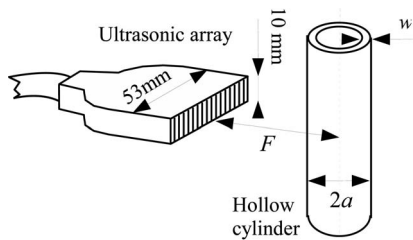
FIG. 1. Experimental setup.

the Taylor expansion of the backscattering matrix. When this expansion is valid, the singular vectors correspond to Legendre polynomials. The validity ranges of the two models are discussed in Sec. VI.

II. EXPERIMENT

A. Setup

Two experimental configurations have been studied. In both cases a linear array of 124 tranducers is used. Each transducer is 0.387 mm wide and the array pitch is 0.417 mm. Hence the size of the array D equals 53 mm. The central frequency is 3 MHz with 60% bandwidth. The central wavelength λ is about 0.5 mm long. As the transducers are high and parallel to the cylinder axis, the wave propagation is almost two-dimensional. The two configurations differ in the distance F between the array and the cylinder, and the cylinder radius a.

A DORT experiment begins with measuring the interelement matrix of an active array of N transducers (see Fig. 1). A pulse is emitted from transducer i. The back-scattered wave is then recorded with the N transducers of the same array. The operation is repeated for the N emitting transducers. The response from transducer i to transducer j is written $k_{ij}(t)$. The $N \times N$ array response matrix $\mathbf{K}(\omega)$ is the Fourier
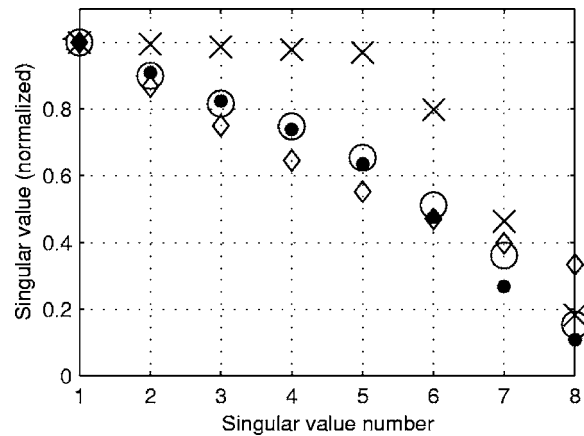
FIG. 2. The eight largest singular values at the central frequency (3 MHz) σ_n of the WR cylinder obtained from experimental data ($\bigcirc$), a cylindrical shell model (Sec. III) without transducer directivity ($\times$) and with transducer directivity ($\bullet$), and a model based on the transducer directivity and the Hermite-Gaussian modes ($\diamond$) [Eq. (20)].

transform of the N^2 impulse responses $k_{ij}(t)$ at the frequency ω. The SVD of the array response matrix $\mathbf{K}$ is then calculated with Matlab.

B. Well-resolved cylinder

The first set of experimental data has been obtained by Komilikis et al. (See Figs. 2 and 3).[15] The reflector consists of a steel 0.5-mm-thick 10-mm-radius air-filled cylinder. The distance F equals 70 mm (Fig. 1). The modulus of the corresponding experimental singular vectors $\mathbf{V}_n$ are presented in Fig. 3. In such a hollow cylinder, circumferential waves (Lamb waves) can propagate. However, the radiation of these waves is weak compared to the specular echo. This is why the singular vectors associated with Lamb waves only appear for rank higher than 8.[15] The analysis of this contribution is provided by Prada and Fink.[16]
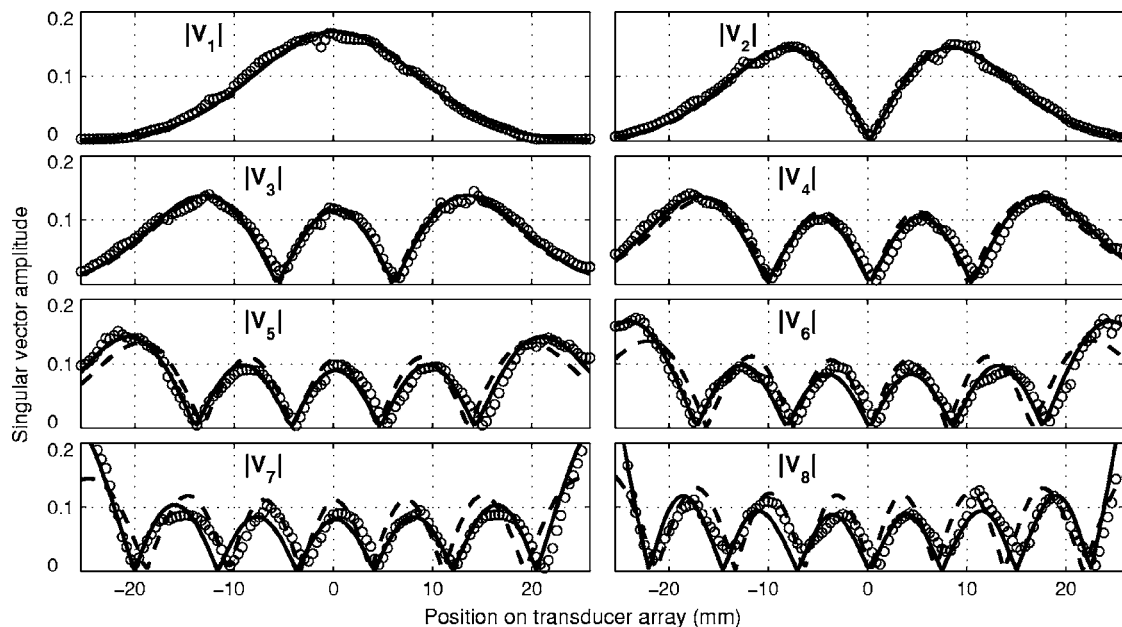
FIG. 3. The eight singular vectors $\mathbf{V}_n$ at the central frequency (3 MHz) of the WR cylinder associated to the eight largest singular values obtained from experimental data ($\bigcirc$), a cylindrical shell model (continuous line) described in Sec. III, and Hermite-Gaussian modes (dashed line) determined in Sec. IV.
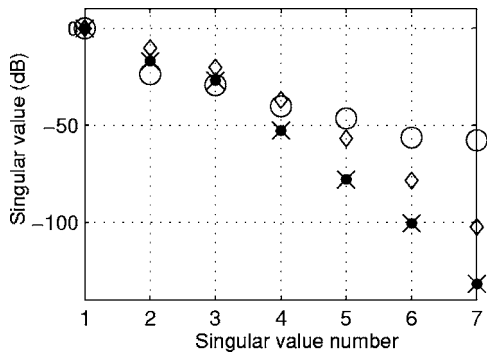
FIG. 4. The seven largest singular values σ_n at the central frequency (3 MHz) of the PR cylinder obtained from experimental data (○), a cylindrical shell model (Sec. III) without transducer directivity (×) and with transducer directivity (●), and a model based on Legendre polynomials (◇) [Eq. (25)].

C. Poorly resolved cylinder

The cylinder radius is much smaller ($a=1.5$ mm) in the second configuration. Its thickness e equals 0.5 mm. Moreover, the cylinder is farther from the array ($F=200$ mm). The seven largest singular values are plotted in Fig. 4. The modulus of singular vectors $\mathbf{V}_n$ corresponding to the six largest singular values are presented in Fig. 5.

III. NUMERICAL CALCULATION OF THE SINGULAR DECOMPOSITION

In order to confirm the experimental results, the backscattering matrix $\mathbf{K}$ has also been numerically computed from an elastic model. At one frequency, the general expression of a pressure field P_{sca} emitted by an isotropic 2D source and scattered by a cylindrical object is given by[17,18]

$$P_{sca}(\mathbf{r}_1, \mathbf{r}_2) = \frac{P_0}{4i} \sum_{l=0}^{\infty} b_l \epsilon_l H_l^{(1)}(k_0 \|\mathbf{r}_1\|) H_l^{(1)}(k_0 \|\mathbf{r}_2\|) \cos(l\theta),$$

(1)

where k_0 is the wave number in water, $H_l^{(1)}$ is the Hankel function of first kind, and P_0 is the incident wave amplitude. $\mathbf{r}_1$ and $\mathbf{r}_2$ are respectively the position of the source and the
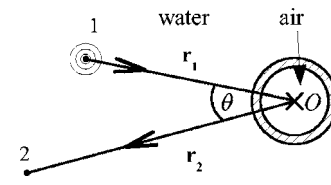
FIG. 6. Schematic view of the quantities that are involved in Eq. (1).

receiver with respect to the center of the object O. θ is the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$ (Fig. 6). $\epsilon_n=1$ when $n=0$ and 2 otherwise. In the case of an elastic cylindrical shell, b_l coefficients are given in Refs. 19 and 20. They depend on the outer and inner radius, the longitudinal and transverse wave speeds, the density of the shell, and finally on the inner and outer fluids.

First, the transducers of the array have been considered as omnidirectional, i.e., $K_{i,j}=P_{sca}(\mathbf{r}_i, \mathbf{r}_j)$ where $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of the center of the ith and jth transducers on the array. The singular vectors of $\mathbf{K}$ fit the experimental ones (Figs. 3 and 5) correctly. The experimental and numerical singular values differ significantly for the WR cylinder (see Fig. 2). In such a case, the transducer directivity cannot be neglected. The transducer directivity has been taken into account by convolving P_{sca} with the aperture function of the array transducers in emission and in reception. The aperture function is assumed to be a square of width e. The interelement matrix that includes the directivity effect is written

$$K_{i,j} = \int_{-e/2}^{e/2} \int_{-e/2}^{e/2} P_{sca}(\mathbf{r}_i - \mathbf{x}_1, \mathbf{r}_j - \mathbf{x}_2) d\mathbf{x}_1 \, d\mathbf{x}_2.$$

(2)

The integrations over $\mathbf{x}_1$ and $\mathbf{x}_2$ are performed along a line that is parallel to the array. In the experiment, $e=1.5\lambda$. e is larger than the array pitch because mechanical coupling between neighbor transducers has to be taken into account in the effective aperture.

The singular values obtained considering directivity are plotted in Figs. 2 and 4. The directivity has an effect only in the case of the WR cylinder: At 70 mm, the 20° array aper-
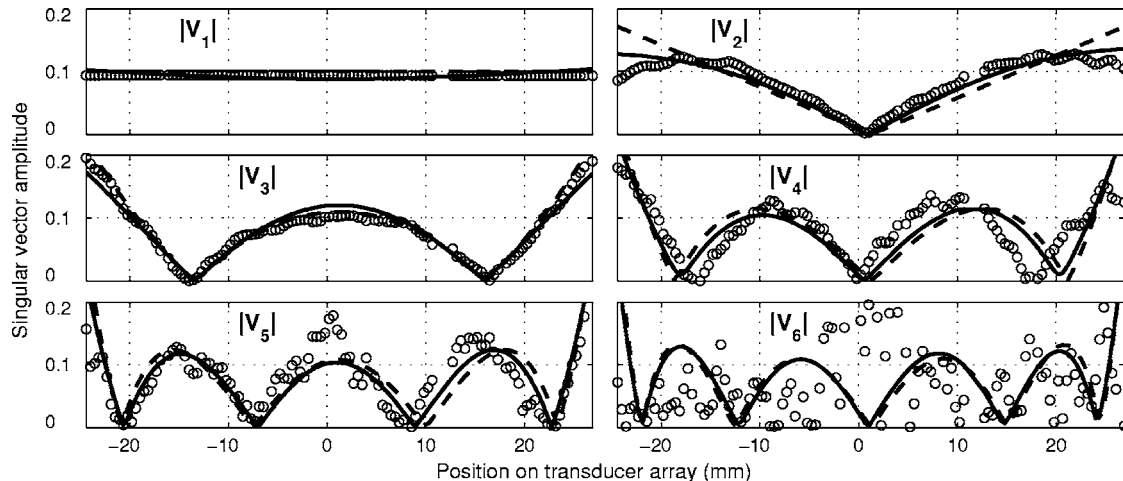
FIG. 5. The six singular vectors $\mathbf{V}_n$ at the central frequency (3 MHz) of the PR cylinder associated to the six largest singular values obtained from experimental data (○), a cylindrical shell model described in Sec. III (continuous line), and Legendre polynomials (dashed line) as in (24).
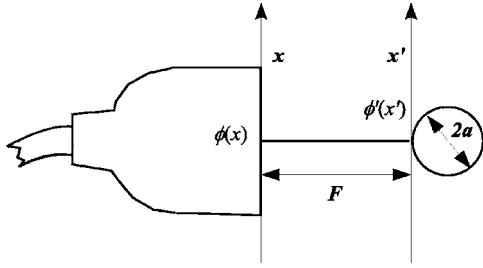
FIG. 7. Schematic top view of the experiment.

ture is comparable to transducer directivity. But at a distance of 200 mm, the angular aperture of the array is three times smaller and the directivity can be neglected.

Although this numerical computation gives exact numerical results, there is no simple analytical expression. Indeed, as the size of the object is large compared to the wavelength, more than 130 terms in the sum contribute to the scattered pressure P_{sca}. Nevertheless, the dominant singular vectors exhibit quite a simple shape for the two configurations. The next two sections show that depending on the value of a and $\sqrt{\lambda a}$ compared to the array resolution $\lambda F/D$, they correspond to Hermite-Gaussian modes or Legendre polynomials.

IV. WELL RESOLVED CYLINDERS: THE HERMITE-GAUSSIAN MODES

A. SVD of continuous fields

A continuous and infinite array is first considered to obtain an analytical expression for the singular vectors and values. This continuous array is at distance F from the edge of the cylinder of radius a (Fig. 7). In the paraxial approximation, the pressure field $\phi'(x')$ on the tangent of the cylinder is given by the Fresnel transformation of the pressure field $\phi(x)$ on the plane of the linear array. With the system being approximately 2D, the field on the tangent becomes

$$\phi'(x') = \int \mathcal{H}_F(x',x)\phi(x)dx \equiv \mathcal{H}_F \times \phi. \qquad (3)$$

$\mathcal{H}_F$ is the kernel of the Fresnel transformation and it is given by

$$\mathcal{H}_F(x',x) = -j\sqrt{\frac{k}{2\pi F}}e^{jkF}e^{j(k/2F)(x-x')^2}, \qquad (4)$$

where k is the wave number and F the distance between the two planes. Using the kernel formalism, the field backscattered toward the array can be worked out. The propagation is composed of three steps: the propagation from the array to the cylinder, the reflection from the cylinder, and then the back propagation from the cylinder to the array. In terms of kernel transformation, the backscattered field ϕ_s is written

$$\phi_s = \mathcal{H}_F \times \mathcal{R}_a \times \mathcal{H}_F \times \phi = \mathcal{K} \times \phi, \qquad (5)$$

where the kernel $\mathcal{R}_a$ describes the specular reflection of the wave from the surface of the cylinder of radius a. The field backscattered towards the array comes from wave reflections on cylinder surface elements located at $x' \ll a$. The parabolic

surface assumption for the cylinder can thus be made. As the cylinder is rigid, only the specular reflection is considered and $\mathcal{R}_a$ is given by

$$\mathcal{R}_a(x_1,x_2) = \mathcal{C}\delta(x_1 - x_2)\exp(-jkx_1^2/a), \qquad (6)$$

where $\mathcal{C}$ is a reflection coefficient of the cylinder and $\exp(-jkx_1^2/a)$ represents the phase delay applied to the backscattered wave field just after reflection due to the cylinder curvature. Thus $\mathcal{H}_F \times \mathcal{R}_a \times \mathcal{H}_F$ represents the scattering kernel $\mathcal{K}$. Performing the SVD means finding the set of singular functions $u_n(x_1)$ and $v_n(x_2)$ associated to the singular value σ_n such as $\mathcal{K}(x_1,x_2) = \Sigma_n u_n(x_1)\sigma_n v_n^*(x_2)$. As the propagation medium is reciprocal, the backscattering kernel is symmetrical, i.e., $\mathcal{K}(x_1,x_2) = \mathcal{K}(x_2,x_1)$. Assuming that the singular values are not degenerated, the reciprocity principle states that $u_n(x) = e^{j\psi}v_n^*(x)$ where ψ is an arbitrary phase. Here, for simplicity, ψ is chosen equal to 0. Finally, the functions v_n verify

$$\sigma_n v_n^* = \mathcal{H}_F \times \mathcal{R}_a \times \mathcal{H}_F \times v_n. \qquad (7)$$

B. Hermite-Gaussian modes and complex radius of curvature

The approach to solve (7) is very similar to the one used to compute the modes of a laser resonator.[21] Indeed, in laser mode theory, the modes of a cavity made of two spherical mirrors are the Gaussian beams that are invariant after reflection from the two mirrors. The Gaussian beams are well adapted to this problem because the wavefronts are parabolic, which corresponds to the parabolic approximations of spherical surfaces. In a 2D problem, the Hermite-Gaussian modes expression is

$$v_n(x) = \sqrt{\frac{2}{\pi w^2}}\xi_{n-1}^H\left(\frac{\sqrt{2}x}{w}\right)\exp\left(-\frac{jkx^2}{2q}\right), \qquad (8)$$

where ξ_n^H are Hermite polynomials and q is the complex radius of curvature (CRC) of the beam at position x. The CRC q is linked to the width w and the radius of curvature of the beam R through the relation

$$\frac{1}{q} = \frac{1}{R} - j\frac{\lambda}{\pi w^2}. \qquad (9)$$

The CRC q is determined from ray transfer matrices,[21] which provide concise and useful representations of both the geometrical propagation of paraxial rays and the propagation with diffraction of Gaussian beams. Hence a simple transformation f changes CRC q_{in} into q_{out} using the expression

$$q_{out} = \frac{Aq_{in} + B}{Cq_{in} + D} \equiv [f] \cdot q, \qquad (10)$$

where A, B, C, and D are the four coefficients of the ray transfer matrix associated with f transformation. The ray transfer matrices, h_F and r_a, corresponding respectively to a propagation over a distance F and the reflection from a cylinder of radius a, are given in Table I.

TABLE I. Ray transfer matrix representation of two Gaussian beam transformations: propagation over a distance F (h_F) and reflection from a cylinder of radius (r_a).

Transformation	h_F	r_a
$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$	$\begin{pmatrix} 1 & F \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 2/a & 1 \end{pmatrix}$

C. Determination of the singular complex radius of curvature

The conjugate of the phase that appears on the left-hand side of Eq. (7) means that the Gaussian beam width does not change but the radius of curvature is of opposite sign. Consequently, for a beam of CRC q, the CRC of the conjugate beam is written $-q^*$. Finally, the ray transfer matrix formulation of Eq. (7) is

$$-q^* = [h_F] \cdot [r_a] \cdot [h_F] \cdot q. \tag{11}$$

Multiplying the two sides of the last equation by $[h_{-F}]$ and using the property $-([h_{-F}] \cdot q)^* = [h_F] \cdot (-q)^*$ (see Appendix B) it follows that

$$-([h_F] \cdot q)^* = [r_a] \cdot [h_F] \cdot q. \tag{12}$$

Introducing the CRC on the cylinder $q' = [h_F] \cdot q$, the following expression is obtained:

$$-q'^* = [r_a] \cdot q'. \tag{13}$$

Equation (13) implies both wave phase and amplitude remain unchanged after reflection from the cylinder, despite the inversion of the propagation direction. Hence, working out the singular vectors on the array is equivalent to working out the invariants by a reflection from the cylinder. As shown in Appendix C, (13) implies that the beam phase matches the shape of the cylinder. In other words, the CRC at the position of the cylinder is given by

$$\frac{1}{q'} = \frac{-1}{a} - j\frac{\lambda}{\pi w'^2}. \tag{14}$$

There is not yet a condition on the width w'. From the relation $q' = [h_F] \cdot q$, i.e., $q = [h_{-F}] \cdot q'$ the beam width at the array is deduced

$$w^2 = w'^2(1 + F/a)^2 + \frac{\lambda^2 F^2}{\pi^2 w'^2}. \tag{15}$$

The square of the singular value is proportional to the energy of the signal that is backscattered within the array aperture. Under the paraxial approximation, the assumption that all the energy is reflected back to the array can be made. Consequently $\sigma_n = 1$ for every singular vector because energy is conserved. In fact, the degeneracy of $\mathcal{K}$ is broken by considering the effect of the transducers' directivity and the array aperture. Taking into account these two effects implies that the invariants may differ from Gaussian modes. However, the directivity and aperture effects can be considered as small perturbations. From perturbation theory, the first order of the singular vectors is given by the singular vectors of the unperturbed system [here $v_n(x)$] while the perturbed singular values are given by

$$\sigma_n = \int_{-D/2}^{D/2} v_n(x)[\mathcal{K}_d \times v_n(x)]dx, \tag{16}$$

where D is the length of the linear array and $\mathcal{K}_d$ is the backscattering kernel that takes into account the directivity of the transducers. Indeed, Gaussian beams of large width are more sensitive to the finite size of the array. Moreover, they involve larger incident angles on the cylinder, which means that the effect of transducer directivity is stronger. As a result, the maximization of singular values leads to the minimization of the Gaussian mode width w. At the minimum w, dw^2/dw'^2 is null. Using (15) yields

$$w'^2 = \frac{\lambda F a}{\pi(F + a)}. \tag{17}$$

From this value, the width and radius of curvature on the array can be deduced:

$$w^2 = 2F\left(1 + \frac{F}{a}\right)\frac{\lambda}{\pi}, \tag{18}$$

$$R = \frac{-2F(1 + F/a)}{1 + 2F/a}. \tag{19}$$

So, the Gaussian beam at the array is uniquely determined.

D. Determination of singular values

Now the dependence of the singular values with respect to the order of Hermite-Gaussian mode n is investigated. The complete computation of σ_n is performed in Appendix D. Assuming small array elements of effective width e, it follows that

$$\sigma_n \propto \int_{-D/2}^{D/2} |v_n(x)\text{sinc}(\pi e/\lambda R)|^2 \, dx. \tag{20}$$

This equation confirms that the order of the Hermito-Gaussian mode corresponds to the order of the singular vector.

In summary, the singular functions of $\mathcal{K}$ are Hermite-Gaussian modes of width and radius of curvature respectively given by (18) and (19). Assuming small size transducers, the singular values are given by (20).

E. Validity of the approximation

There are two requirements on the Gaussian beams to properly fit at least the first singular function. First, the beam width at the cylinder position should be smaller than the cylinder radius ($w' < a$) to carry out the local approximation of a parabolic shape. Second, in order to observe the complete Gaussian beam on the array, the Gaussian width at the array position should be smaller than the array size ($w < D$). From (17) and (18), it follows that

$$\lambda < a \tag{21}$$

and

Aubry *et al.*: Gaussian and Legendre invariants

$$\lambda F/D < \sqrt{a\lambda}. \tag{22}$$

If these conditions are fulfilled, singular vectors correspond to Hermite-Gaussian modes.

V. POORLY RESOLVED CYLINDERS: THE LEGENDRE POLYNOMIALS

A. Legendre polynomials as singular vectors

The b_l coefficients in (1) are null when l becomes larger than ka. Consequently, the $\cos(l\theta)$ terms are monotonic when $ka\theta < 1$. Since the value of θ is bounded by D/F, the backscattered field smoothly depends on emitting/receiving positions on the array when $\lambda F/D < a$. If this last condition is fulfilled, the backscattered pressure field can be expressed with the following Taylor expansion,

$$P_{sca}(x_1,x_2) = e^{ik(r_1+r_2)} \sum_{n_1,n_2} \frac{\partial_{x_1}^{n_1}\partial_{x_2}^{n_2}\hat{P}_{sca}(0,0)}{n_1!n_2!} x_1^{n_1}x_2^{n_2}, \tag{23}$$

where $P_{sca}(x_1,x_2) = e^{ik(r_1+r_2)}\hat{P}_{sca}(x_1,x_2)$, $r_1^2 = F^2 + x_1^2$, and $r_2^2 = F^2 + x_2^2$. In Appendix A, it is shown that if the Taylor coefficients decrease sufficiently rapidly, the eigenfunctions are given by

$$v_n(x) \approx e^{ikr}\widetilde{\mathcal{L}}_{n-1}(x), \tag{24}$$

and the singular values by

$$\sigma_n \approx D\left(\frac{D}{2F}\right)^{2n}\frac{(2n+1)^2}{(2n+1)!!^2}\partial_{\chi_1}^n\partial_{\chi_2}^n Q(0,0), \tag{25}$$

where $\widetilde{\mathcal{L}}_n$ is the normalized Legendre polynomial of order n (see Ref. 22) and $Q(\chi_1,\chi_2) = \hat{P}_{sca}(\chi_1 F, \chi_2 F)$. Therefore the singular functions v_n are proportional to the Legendre polynomials $\widetilde{\mathcal{L}}_{n-1}$ of order $n-1$. When $F \gg D$, this result is in agreement with Zhao's ones.[14] Zhao works out the first two singular vectors and values from a Taylor expansion of the backscattered field from a flat target. Zhao's model seems to work for a larger range of distances F, but its analysis is restricted to the first two singular vectors and values. The Legendre polynomials model is less accurate but provides the whole basis of singular vectors.

B. Validity of the approximation

In order to observe the Legendre polynomials as singular vectors, the Taylor coefficients should decrease as the indexes increase. This usually happens when the function is sufficiently monotonic and smooth. For the cylinder, the backscattered field on the array slowly fluctuates when $\lambda F/D$ is small compared to the cylinder radius a. However, for more complex scatterers (resonant scatterers for instance), different criteria that involve parameters other than the typical size of the scatterer may have to be introduced. In such PR regime, the singular values decrease very quickly with n. The first obvious consequence is the difficulty to measure them experimentally. The second consequence is the extreme sensitivity of the singular values to a small mismatch between the experimental and theoretical singular vectors. For instance, one can assume a weak coupling between the first

and the sixth theoretical singular vectors. The sixth experimental singular vector is for instance given by $\widetilde{\mathcal{L}}_5 + \widetilde{\mathcal{L}}_0/10$. The singular value equals now $\beta_{0,0} + \beta_{5,5}/100 + \beta_{1,5}/10 + \beta_{5,1}/10$, but $\beta_{0,0}$ is $(F/D)^5$ larger than $\beta_{5,5}$, consequently the sixth singular value can be completely modified (β coefficients are given in Appendix A). Hence in the PR approximation, even if the singular vectors seem to be well-fitted by Legendre polynomials, a small mismatch between theoretical and experimental singular vectors implies a strong deviation of singular values from the theoretical prediction given in (25).

VI. DISCUSSION

In the experiment, the array is discrete and the SVD is applied to the response matrix $\mathbf{K}$. The singular vectors $\mathbf{V}_n$ correspond to singular functions v_n given by (8) and (24) evaluated at each transducer position x_j:

$$\mathbf{V}_n = [v_n(x_1) \quad v_n(x_2) \quad \cdots \quad v_n(x_N)]. \tag{26}$$

The comparison between analytical, experimental, and numerical singular vectors and values are shown in Figs. 2 and 3 for the WR cylinder and Figs. 4 and 5 for the PR one.

A. Well-resolved cylinder

Numerical calculations of the singular vectors give the same results with or without considering transducer directivity and are in a good agreement with experiment (Fig. 3). On the contrary, the singular values depend mainly on the directivity of the transducers (Fig. 2), so the use of the perturbation theory is judicious. An excellent agreement between numerical calculation considering transducer directivity, Hermite-Gaussian modes, and experiment is found for the first five singular values and vectors (Figs. 2 and 3).

A numerical simulation considering omnidirectional transducers shows that for $n \geq 6$, the singular values decrease because of the finite size of the array. In a similar way, for $n \geq 6$, the mismatch between the Hermite-Gaussian modes is especially noticeable on the edges of the array. So, beyond this order, the finite size of the array cannot be considered as a small perturbation and the singular vectors are not simply given by Hermite-Gaussian modes. The typical position of the last maximum of a Hermite-Gaussian mode of order n is $w\sqrt{(2n-1)/2}$ where w is given by (19). Here $w = 11.6$ mm, thus the last maximum of the Hermite-Gaussian mode reaches the edge of the array when $n = 9$, which is roughly in agreement with the observed value 6. Above this order, the Hermite-Gaussian modes no longer correspond to the singular vectors. This criterion gives a qualitative limit for the analytical model.

This model assumes a specular echo. Strictly speaking, it could only be applied to rigid cylinders. However, in the case of elastic targets, if the specular energy is superior to elastic wave radiation, Gaussian beam theory can still be applied. The specular energy lost due to elastic waves can be modeled by introducing a reflection coefficient $\mathcal{C} < 1$ in (6).

B. Poorly resolved cylinder

The Legendre polynomials fit the first five singular vectors (Fig. 5) quite well. Beyond that, the singular vectors do not show coherent structure. They are dominated by experimental uncertainties such as electronic noise or sensitivity fluctuations between the transducers of the array. For the same reason the experimental singular values seem to reach a −60-dB plateau for $n > 5$. The mismatch between the experimental/numerical/theoretical singular values is significant (Fig. 4). As explained before, the singular values are extremely sensitive to a small deviation of the singular vectors. Nevertheless (25) is useful to provide an overall trend. It has been checked numerically that σ_n decreases as $(D/2F)^{2n}$ with the distance between the array and the transducer. Chambers and Gautesen's numerical results confirm the role played by Legendre polynomials in a more complex configuration.[10] The first seven singular vectors are computed for a bubble and a rubber sphere with radius comparable to or larger than the wavelength. In both cases, the singular vectors look like Legendre polynomials. Hence the correspondence between Legendre polynomials and singular vectors seems to be a more general concept than the restricted quasi-two-dimensional configuration. This model does not assume that the cylinder is rigid. The results may thus be generalized to soft targets, contrary to the Gaussian beam model.

VII. CONCLUSION

In this article, the singular value decomposition of the backscattering matrix for a rigid cylinder large with respect to the wavelength has been studied. On one hand, when the array size is small ($\lambda F/D > a$), Legendre polynomials fit the singular vectors well. These Legendre polynomials spread out all over the array. Only the singular values depend on the cylinder's physical parameters. On the other hand, when the array is large ($\lambda F/D < \sqrt{\lambda a}$), Hermite-Gaussian modes are a good approximation of the singular vectors. The width of modes depends on the cylinder radius while the singular values are mostly sensitive to the array directivity. Hence a direct application of this study would be the determination of the cylinder radius from the singular vector behavior in the case of a well-resolved cylinder. The methodology used here to compute the singular vectors of the specular echo that is reflected by a large cylinder could easily be generalized to many other large objects with paraboliclike surface.

APPENDIX A: TAYLOR EXPANSION AND LEGENDRE FUNCTIONS

The Taylor expansion of the backscattered pressure field is written

$$\hat{P}_{sca}(x_1,x_2) = \sum_{n_1,n_2} \frac{\partial_{x_1}^{n_1}\partial_{x_2}^{n_2}\hat{P}_{sca}(0,0)}{n_1!n_2!}x_1^{n_1}x_2^{n_2}, \tag{A1}$$

where $\hat{P}_{sca}(x_1,x_2)$ is the backscattered pressure field on a linear array. x_1 and x_2 are respectively the positions of the emitter and the receiver. The distance between the array and the target F is taken into account by introducing a function Q_{sca} such that $Q_{sca}(\chi_1,\chi_2) = \hat{P}_{sca}(\chi_1 F,\chi_2 F)$. Moreover the positions x_1 and x_2 are normalized by half the size of the array (D). It follows that

$$\hat{P}_{sca}(x_1,x_2) = \sum_{n_1,n_2} \left(\frac{D}{2F}\right)^{n_1+n_2}\frac{\partial_{\chi_1}^{n_1}\partial_{\chi_2}^{n_2}Q(0,0)}{n_1!n_2!}\left(\frac{2x_1}{D}\right)^n\left(\frac{2x_2}{D}\right)^n. \tag{A2}$$

An nth power can be expressed in terms of Legendre polynomials[22]

$$x^n = \sum_{l=n,n-2,n-4,\ldots} \frac{(2l+1)n!}{2^{(n-l)/2}[(n-l)/2]!(l+n+1)!!}\mathcal{L}_1(x), \tag{A3}$$

where $\mathcal{L}_l$ is the lth-order Legendre function. The normalized Legendre function ($\tilde{\mathcal{L}}_l$) is introduced. It is defined by

$$\tilde{\mathcal{L}}_l(x) = \sqrt{\frac{2l+1}{D}}\mathcal{L}_l\left(\frac{2x}{D}\right). \tag{A4}$$

Due to the Legendre polynomial properties, the normalized Legendre polynomials define a orthonormal basis ($\int_{-D/2}^{D/2}\tilde{\mathcal{L}}_l\tilde{\mathcal{L}}_l'(x)\,dx = \delta_{l,l'}$, where δ is the Kronecker symbol).

Combining (A2)–(A4), it follows that

$$\hat{P}_{sca}(x_1,x_2) = \sum_{l_1,l_2} \tilde{\mathcal{L}}_{l_1}(x_1)\tilde{\mathcal{L}}_{l_2}(x_2)\beta_{l_1,l_2}, \tag{A5}$$

where

$$\beta_{l_1,l_2} = D\sum_{n_1=l_1,l_1+2,\ldots}\sum_{n_2=l_2,l_2+2,\ldots}\left(\frac{D}{2F}\right)^{n_1+n_2}\partial_{\chi_1}^{n_1}\partial_{\chi_2}^{n_2}Q(0,0)\alpha_{n_1}^{l_1}\alpha_{n_2}^{l_2}. \tag{A6}$$

Coefficients α_n^l are defined as follows:

$$\alpha_n^l = \frac{\sqrt{2l+1}}{2^{(n-l)/2}([n-l]/2)!(l+n+1)!!}. \tag{A7}$$

Assuming that the array aperture is very small, ($D/F \ll 1$), only the first term in (A6) is kept ($l_1 = n_1$ and $l_2 = n_2$). Now β_{l_1,l_2} is approximated by

$$\beta_{l_1,l_2} \approx D\left(\frac{D}{2F}\right)^{l_1+l_2}\frac{(2l_1+1)(2l_2+1)}{(2l_1+1)!!(2l_2+1)!!}\partial_{\chi_1}^{l_1}\partial_{\chi_2}^{l_2}Q(0,0). \tag{A8}$$

Here a qualitative argument is proposed to justify that the Legendre polynomials are the singular functions. The "matrix" made of the β_{n_1,n_2} coefficients is dominated by the $\beta_{0,0}$ element (zeroth order in D/F). Considering only this element, the singular value is given by $\beta_{0,0}$ and the singular vector is $\tilde{\mathcal{L}}_0$. The second singular function can be deduced

from the previous matrix where the first column and the first line are removed because of the orthogonality between the singular functions. Now this new matrix is dominated by $\beta_{1,1}$ element (second order in D/F). Again the singular value is given by $\beta_{1,1}$ and the singular function by $\tilde{\mathcal{L}}_1$. The same argument can be used to work out all the singular values and functions. Finally, the singular vectors are given by

$$v_n(x) \approx e^{ikr}\tilde{\mathcal{L}}_n(x), \tag{A9}$$

and the singular values by

$$\sigma_n \approx D\left(\frac{D}{2F}\right)^{2n}\frac{(2n+1)^2}{(2n+1)!!^2}\partial_{\chi_1}^n\partial_{\chi_2}^n Q(0,0). \tag{A10}$$

APPENDIX B: TRANSLATION OF CONJUGATE GAUSSIAN BEAMS

Here it is shown that the translation by a distance F of a phase conjugate CRC q is equal to the phase conjugate of translated CRC q by a distance $-F$. Indeed,

$$[h_F]\cdot(-q^*) = -q^* + F = -(q^* - F) = -(q - F)^*$$
$$= -([h_{-F}]\cdot q)^*. \tag{B1}$$

The last equation corresponds well to the initial statement.

APPENDIX C: COMPLEX RADIUS OF CURVATURE REFLECTION

The time-reversal invariance of the Gaussian beam after reflection from a cylinder of radius a is written

$$-q'^* = [r_a]\cdot q', \tag{C1}$$

where q' is the CRC of the Gaussian beam. Using the matrix elements of the reflection transformation (see Table I), it follows that

$$-q'^* = \frac{q'}{2q'/a+1}. \tag{C2}$$

This relation is easier to interpret in terms of inverse of CRC:

$$\mathrm{Re}\left(\frac{1}{q'}\right) = -\frac{1}{a}. \tag{C3}$$

The last equation implies that the radius of curvature of the beam equals $-a$ at the position of the cylinder. Note that the reflection invariance does not set a condition on the beam width.

APPENDIX D: PHASE CONJUGATE INVARIANT

The finite size of the transducer acts as a low-pass filter in emission and in reception that eliminates the high spatial frequencies. Formally, the directivity is taken into account by introducing a smoothing function $W(x/e)$, where W is the square function[23] and e is the effective size of one transducer on the array. The perturbed kernel $\mathcal{K}_p$ is written

$$\mathcal{K}_p = \mathcal{D}\times\mathcal{K}\times\mathcal{D}, \tag{D1}$$

where $\mathcal{D}$ is the directivity function ($\mathcal{D}(x,x') = W[(x-x')/e]$). When v_n is emitted by the array, the backscattered field

$S(x) = \mathcal{K}_p\times v_n(x)$ is then recorded on the array. $S(x)$ is not proportional to v_n^* due to the transducer directivity. Using the convolution operator $\otimes$, $S(x)$ becomes

$$S(x) = \{[\mathcal{H}_F(x)\otimes W(x/e)\otimes v_n(x)]\cdot r_a(x)\}\otimes\mathcal{H}_F(x)$$
$$\otimes W(x/e), \tag{D2}$$

where $\mathcal{H}_F(x)\equiv\mathcal{H}_F(x,0)$ and $\mathcal{R}_a(x,x')\equiv\delta(x-x')r_a(x)$. Introducing the field on the cylinder $v_n'(x)$, it follows that

$$S(x)\propto\{[W(x/e)\otimes v_n'(x)]\cdot r_a(x)\}\otimes\mathcal{H}_F(x)\otimes W(x/e). \tag{D3}$$

To proceed further, e is assumed to be small. More precisely one can show that if $e\ll\sqrt{\lambda a/4\pi}$, the last equation can be approximated by

$$S(x)\propto W(x/e)\otimes[v_n'(x)r_a(x)]\otimes\mathcal{H}_F(x)\otimes W(x/e). \tag{D4}$$

$v_n'(x)$ has been chosen such that $v_n'(x)r_a(x) = v_n'^*(x)$ because it is invariant under reflection from the cylinder. Using this property, it follows that

$$S(x)\propto W(x/e)\otimes v_n'^*(x)\otimes\mathcal{H}_F(x)\otimes W(x/e). \tag{D5}$$

Introducing the conjugate field on the array, the expression is simplified:

$$S(x)\propto W(x/e)\otimes v_n^*(x)\otimes W(x/e). \tag{D6}$$

$v_n(x)$ is a normalized function, so the perturbed singular value is given by

$$\sigma_n = \int_{D/2}^{D/2}S(x)v_n(x)dx \tag{D7}$$

$$= \int_{D/2}^{D/2}|W(x/e)\otimes v_n(x)|^2\,dx. \tag{D8}$$

The finite aperture size D is taken into account at the integration boundaries. The expression can be simplified when $e\ll\sqrt{D\lambda/2\pi}$ and it finally follows that

$$\sigma_n\propto\int|v_n(x)\mathrm{sinc}(\pi e/\lambda R)|^2\,dx. \tag{D9}$$

[1]C. Prada and M. Fink, "Eigenstates of the time reversal operator: A solution to selective focusing in multiple-target media," Wave Motion **20**, 151–163 (1994).

[2]E. Kerbrat, C. Prada, D. Cassereau, and M. Fink, "Imaging in the presence of grain noise using the decomposition of the time reversal operator," J. Acoust. Soc. Am. **113**, 1230–1240 (2003).

[3]J.-L. Robert, M. Burcher, C. Cohen-Bacrie, and M. Fink, "Time reversal operator decomposition with focused transmission and robustness to speckle noise: Application to microcalcification detection," J. Acoust. Soc. Am. **119**, 3848–3859 (2006).

[4]N. Mordant, C. Prada, and M. Fink, "Highly resolved detection and selective focusing in a waveguide using the d.o.r.t method," J. Acoust. Soc. Am. **105**, 2634–2642 (1999).

[5]J. F. Lingevitch, H. C. Song, and W. A. Kuperman, "Time reversed reverberation focusing in a waveguide," J. Acoust. Soc. Am. **111**, 2609–2614 (2002).

[6]C. F. Gaumond, D. M. Fromm, J. F. Lingevitch, R. Menis, G. F. Edelmann, D. C. Calvo, and E. Kim, "Demonstration at sea of the decomposition-of-the-time-reversal-operator technique," J. Acoust. Soc. Am. **119**, 976–990 (2006).

[7]H. Tortel, G. Micolau, and M. Saillard, "Decomposition of the time rever-

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Aubry *et al.*: Gaussian and Legendre invariants 2753

sal operator for electromagnetic scattering," J. Electromagn. Waves Appl. **13**, 687–719 (1999).

[8]G. Micolau, M. Saillard, and P. Borderies, "Dort method as applied to ultrawideband signals for detection of buried objects," IEEE Trans. Geosci. Remote Sens. **41**, 1813–1820 (2003).

[9]S. K. Lehman and A. J. Devaney, "Transmission mode time-reversal super-resolution imaging," J. Acoust. Soc. Am. **113**, 2742–2753 (2003).

[10]D. H. Chambers and A. K. Gautesen, "Time reversal for a single spherical scatterer," J. Acoust. Soc. Am. **109**, 2616–2624 (2001).

[11]J.-G. Minonzio, C. Prada, D. Chambers, D. Clorennec, and M. Fink, "Characterization of subwavelength elastic cylinders with the decomposition of the time-reversal operator: Theory and experiment," J. Acoust. Soc. Am. **117**, 789–798 (2005).

[12]D. H. Chambers and J. G. Berryman, "Analysis of the time-reversal operator for a small spherical scatterer in an electromagnetic field," IEEE Trans. Antennas Propag. **52**, 1729–1738 (2004).

[13]D. Chambers, "Analysis of the time reversal operator for scatterers of finite size," J. Acoust. Soc. Am. **112**, 411–419 (2002).

[14]H. Zhao, "Analysis of the response matrix for an extended target," SIAM J. Appl. Math. **64**, 725–745 (2004).

[15]S. Komilikis, C. Prada, and M. Fink, "Characterization of extended objects with the d.o.r.t. method," IEEE Ultrasonics Symposium **20**, 1401–1404 (1996).

[16]C. Prada and M. Fink, "Separation of interfering acoustic scattered signals using the invariants of the time-reversal operator application to lamb waves characterization," J. Acoust. Soc. Am. **104**, 801–807 (1998).

[17]G. V. Frisk and H. Uberall, "Creeping waves and lateral waves in acoustic scattering by large elastic cylinders," J. Acoust. Soc. Am. **59**, 46–54 (1976).

[18]A. Derem, *La diffusion acoustique* (CEDOCAR, Paris, 1989), Chap. 9, pp. 198–200.

[19]R. D. Doolittle and H. Uberall, "Sound scattering by elastic cylindrical shells," J. Acoust. Soc. Am. **39**, 272–275 (1966).

[20]J.-L. Rousselot, *La diffusion acoustique* (CEDOCAR, Paris, 1989), Chap. 12, p. 351.

[21]C. Webb and J. Jones, *Handbook of Laser Technology and Applications-Volume 1: Principles* (Institute of Physics, Philadelphia, 2004).

[22]E. W. Weisstein, "Legendre polynomial," From MathWorld-A Wolfram Web Resource, http://mathworld.wolfram.com/LegendrePolynomial.html.

[23]$W(x) = 1$ when $|x| < 1/2$ and $W(x) = 0$ otherwise.

2754   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Aubry *et al.*: Gaussian and Legendre invariants

Active waveguide Green's function estimation with application to time-reversal focusing without a probe source in a range-independent waveguide

S. C. Walker[a),b)] and W. A. Kuperman[c)]

Marine Physical Laboratory of the Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238

Philippe Roux[d)]

LGIT- Maison des Géosciences, 1381 rue de la Piscine, BP 53 - 38041 Grenoble cedex 9, France

A method is introduced for using a pair of vertical line transducer arrays (VLAs) to produce time-reversal (TR) foci at any location in a range-independent shallow-ocean waveguide without the need for a probe source. Modal theory suggests that the acoustic response sampled between a pair of VLAs separated by a range R can be repeatedly iterated to estimate the acoustic response at ranges that are integer multiples of R. When combined with the frequency-based variable-range-TR method, the process results in a virtual sampling of the acoustic response over long distances that can be applied to both passive and active shallow-ocean acoustic imaging. Here, the technique is applied to active time-reversal focusing. Array geometry effects are explored and simulation and experimental results are presented. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2355538]

I. INTRODUCTION

The ability of time reversal (TR) to generate high-resolution acoustic focusing in the shallow ocean makes it an attractive candidate for active and passive sonar and communications applications as well as matched-field processing techniques.[1,2] The major limitation of current TR methods is that they require an acoustic probe source at the desired focal location. This work discusses a method for using a pair of vertical line transducer arrays (VLAs) to produce TR foci at distant ranges at any depth in the water column without the need for a probe source.

By Green's theorem, the acoustic response over a closed surface completely determines the response throughout the enclosed volume. In an experimental context, a dense sampling of the point-to-point acoustic response between a pair of vertical transducer arrays separated in range can be used to uniquely determine the response throughout the intervening volume. In the case of a range-independent, static propagation medium, the point-to-point acoustic response is reciprocal and independent of absolute position, depending only on the source and receiver depths and on their relative range separation (throughout the remainder of this paper, the relative range separation between the VLAs is referred to as the iteration range). Modal theory suggests that the point-to-point acoustic response between a pair of VLAs can be iterated to estimate the acoustic response at integer multiples of the iteration range. According to Huygen's principle, the re-

sponse on one array due to the other can be decomposed into wavelets that govern the phase evolution of the field over the iteration range. Iterating the response function is equivalent to evolving the response phase over integer multiples of the iteration range. The iterated response estimates can be used to generate a TR focus at integer multiples of the iteration range at any sampled depth.

In a mathematical context, time reversal is applied as a cross correlation between the sampled VLA field (due to a point source for example) and the Green's function (as a function of range and depth) summed over the array elements. The iteration process, on the other hand, is mathematically applied as a cross convolution of the sampled point-to-point pressure response (between the array pair) summed over one of the arrays. Whereas the correlation process (phase conjugation of the source field) brings the field back into focus back at the source origin, convolution propagates the field further down range. As the convolution is a multiplicative process, repeated iteration can lead to undesirable narrowing of the frequency spectrum.

The virtual sampling generated by the iteration process, termed sampled response iteration (SRI), is limited to integer multiples of the iteration range. Applying the theory of acoustic field invariants in a range-independent waveguide,[3] SRI can be combined with variable-range TR (Ref. 4) to extend the method to intervening ranges, i.e., nonintegral multiples of the iteration range.

Section II presents a development of the theory of the iteration method and accompanying simulation results. A discussion of the experimental setup in Sec. II A is followed in Sec. II B by a theoretical development. Section II B 1 introduces the modal acoustic model. The theory of the iteration method is developed in Sec. II B 2. The theoretical discus-

---

[a)]Electronic mail: shane@physics.ucsd.edu
[b)]Also at Physics Department, University of California, San Diego.
[c)]Electronic mail: wak@mpl.ucsd.edu
[d)]Electronic mail: philippe.roux@obs.ujf-grenoble.fr
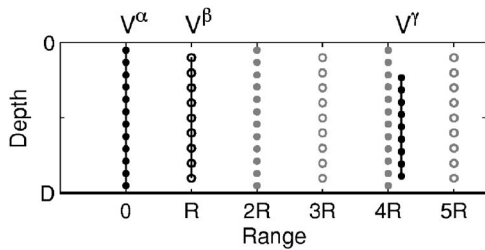
FIG. 1. Experimental setup. The point-to-point acoustic pressure response sampled between transducer arrays V^α (closed circles) and V^β (open circles) is propagated to generate virtual sampling at integer multiples of the array separation, R. The virtual sampling locations are indicated with gray closed and open circles. Time-reversing the virtual field corresponding to one of the virtual samples results in a TR focus that is measured by a third array, V^γ (closed circles).

sion addresses the iteration technique as well as associated bandwidth equalization issues. The effects of array geometry are considered in Sec. II B 3. In Sec. II B 4 the iteration method is extended to the entire waveguide in combination with variable-range TR. Results from ultrasonic experiments in a laboratory waveguide are presented in Sec. III. Section IV gives a summary and concluding remarks.

II. THEORY AND SIMULATION OF THE SAMPLED RESPONSE ITERATION (SRI) PROCEDURE

A. Experimental setup

The range-independent, azimuthally symmetric experimental geometry for which the SRI procedure is designed is schematically represented in Fig. 1. A pair of parallel, vertical-geometry transducer element arrays, labeled V^α (closed circles) and V^β (open circles), are located a distance R apart, where R is the iteration range. A cylindrical coordinate system is chosen such that the origin ($r=0$) coincides with V^α. The point-to-point acoustic response, sampled between all transducer elements of the array pair, is iterated to generate response function estimates at successive intervals of the iteration range. The net result is that the point-to-point response is virtually sampled in range with R as the range sampling interval. The gray closed and open circles of Fig. 1 represent points at which the field is virtually sampled. By applying a frequency shift to the iterated fields, it is possible to extend the virtual sampling to intervening ranges. For the purposes of time reversal, the set of virtual samples represents an ensemble of fields from virtual sources. Time reversal (phase conjugation) and transmission of a chosen virtual source field results in an acoustic focus at the location of the virtual source. In order to simplify the analysis, the arrays are always labeled such that V^α corresponds to the TR field transmission array. As seen in Fig. 1, a third array, labeled V^γ, is used to sample the resulting TR field.

B. Theoretical discussion

After summarizing the relevant modal theory, this section introduces the range iteration method. Array geometry and bandwidth effects are addressed. Finally, the SRI method is combined with the variable-range TR method of Ref. 4. The theoretical analysis is illustrated with simulation results.

1. The modal point-to-point pressure response

The elements of arrays V^α and V^β are labeled by the indices a and b, respectively. The ath element of array V^α is at depth $z=z_a$ and likewise for V^β. The arrays are separated by a distance R. Array V^α comprises a total of A elements while array V^β comprises B elements. Applying the convolution theorem, the ω angular frequency component of the acoustic pressure response at element a on V^α due to a source at element b on V^β is written as the product

$$P_{ab}(R,\omega) = G_{ab}(R,\omega)S(\omega), \tag{1}$$

where $G_{ab}(R,\omega)$ represents the point-to-point Green's function between elements a and b, and $S(\omega)$ represents the ω component of the Fourier decomposition of the time-domain source signal.

The Green's function in terms of modal depth functions, $\phi_m(z,\omega)$, and associated modal wave numbers, $k_m(\omega)$, is written[5]

$$G_{ab}(R,\omega) = \frac{e^{-i(\pi/4)}}{\rho(z_b)\sqrt{8\pi R}} \sum_{m=1}^{M} \frac{\phi_m(z_a,\omega)\phi_m(z_b,\omega)}{\sqrt{k_m(\omega)}} e^{-ik_m(\omega)R},$$
$$\tag{2}$$

with associated closure and orthogonality relations,

$$\sum_{m=1}^{\infty} \frac{\phi_m(z,\omega)\phi_m(z',\omega)}{\rho(z')} = \delta(z-z') \tag{3}$$

$$\int_0^D \frac{\phi_m(z,\omega)\phi_n(z,\omega)}{\rho(z)} dz = \delta_{mn}. \tag{4}$$

Here, $z=0$ corresponds to the water surface and $z=D$ is the depth spanned by the modal depth functions, while $\rho(z)$ represents the depth-dependent mass density.

2. SRI theory

a. The singly iterated field. Convolution of a signal by the Green's function generates a forward propagation through the medium in positive time. This suggests the possibility of virtually propagating the pressure response between a pair of arrays to further ranges through iterative convolution. To demonstrate, suppose the pressure response on array pair V^α/V^β has been determined through measurement [Eq. (1)]. Taking the pressure response, $P_{ba'}(R,\omega)$, on V^β due to the source at element a' on V^α as a collection of source fields (invoking reciprocity to switch the indices), the propagation operation is defined as the cross convolution of each source field by the full point-to-point pressure response, $P_{ab}(R,\omega)$, summed over the elements of V^β (note that $[P_{ba'}]$ is a vector over the elements of V^β while $[P_{ba}]$ is a two-dimensional matrix over the elements of both arrays). Using Eqs. (1) and (2), the resulting virtually propagated field, $\hat{P}(\omega)$, is

$$\hat{P}_{aa'}(\omega) = \sum_{b=1}^{B} P_{ab}(R,\omega) P_{ba'}(R,\omega) \Delta z_b$$

$$= \frac{e^{-i\pi/2} S^2(\omega)}{8\pi\rho(z_{a'})} \sum_{m,n=1}^{M,M} \frac{\phi_m(z_a,\omega)\phi_n(z_{a'},\omega)}{\sqrt{k_m(\omega)R}\sqrt{k_n(\omega)R}}$$

$$\times e^{-i(k_m(\omega)+k_n(\omega))R} \left(\sum_{b=1}^{B} \frac{\phi_m(z_b,\omega)\phi_n(z_b,\omega)}{\rho(z_b)} \Delta z_b \right).$$

(5)

The quantity Δz_b represents the interelement spacing associated with the transducer element at $z = z_b$. Applying the mode orthogonality relation of Eq. (4) to the term in parentheses in the limit where V^β densely samples most of the water column, Eq. (5) simplifies to

$$\hat{P}_{aa'}(\omega) \approx \frac{e^{-i\pi/2} S^2(\omega)}{8\pi\rho(z_{a'})} \sum_{m=1}^{M} \frac{\phi_m(z_a,\omega)\phi_m(z_{a'},\omega)}{k_m(\omega)R} e^{-i2k_m(\omega)R}$$

$$\approx e^{-i\pi/4} S(\omega) \left(\sqrt{\frac{1}{4\pi R}} \right) P_{aa'}(2R,\omega).$$

(6)

Orthogonality condition (4) is valid for every mode for which the array geometry samples at two points per period (in depth) over the entire effective span of the mode. Typically, an array sampling interval equal to or less than half the characteristic acoustic wavelength of the water column is sufficient to adequately sample all the modes.

Ignoring for the moment the factor in Eq. (6), $e^{-i\pi/4} S(\omega)(\sqrt{1/4\pi R})$, the virtual propagation result closely approximates the frequency response between array V^α and a second virtual array at range $2R$. Effectively, the procedure has propagated the field an additional range increment R. In the case of a single iteration, the virtual array is of the same geometric configuration as V^α. As will be shown, the geometry of the final virtual array alternates between the geometries V^α or V^β with each iteration.

Although the iteration procedure preserves the relative phases of the modal components, it produces errors associated with cylindrical spreading. As a result, in deriving the approximation of Eq. (6), a factor of $1/\sqrt{k_m(\omega)}$ has been neglected for each modal component in proceeding from the first line to the second line. This is justifiable because the modal wave numbers typically do not differ significantly. For example, under realistic shallow-ocean conditions, the modal wave numbers differ at most by a factor of 2 between the highest and lowest order modes. Indeed, the experimental success of the iteration method validates the theoretical assumption that the mode dependence of the cylindrical spreading factor is negligible.

In both simulation and experiment, the collection of virtual responses can be time-reversed and broadcast from array V^α to produce an acoustic focus at all sampled depths at range $2R$. Repeated iteration results in the virtual responses at multiple integers of R.

b. The multiply iterated field. Ignoring for the moment the practical considerations involved with deploying a pair of parallel, full-spanning, densely sampling arrays, this result implies that the field can be virtually sampled over long ranges without any knowledge of the environment by simply iterating this procedure. To demonstrate this, it is convenient to identify the acoustic response between the arrays as a matrix,

$$\mathbf{P} \equiv [P_{ab}(R,\omega)].$$

(7)

For simplicity, the explicit dependence on R and ω is suppressed in the matrix notation.

In order to be as general as possible, the iteration method is developed for the case of arrays of differing geometries comprising different numbers of elements ($A \neq B$). In this case $\mathbf{P}$ is not a square matrix. Depending on the implementation and number of iterations, the iterated, virtually sampled field, $\hat{\mathbf{P}}$, may be an $A \times A$, $A \times B$, $B \times A$, or $B \times B$ matrix. In what follows, the iteration procedure is developed to yield virtually sampled fields that facilitate TR imaging from array V^α, i.e., the final virtually sampled response always has A components.

The multiply iterated virtually propagated response is given as

$$\hat{\mathbf{P}}_{\mathbf{N}} = \mathbf{Q}_{\mathbf{N}} \mathbf{Q}_{\mathbf{0}},$$

(8)

where $\mathbf{Q}_{\mathbf{0}}$ is the starter field, N is the number of applied iterations, and $\mathbf{Q}_{\mathbf{N}}$ is the iterated field. The starter field is defined

$$\mathbf{Q}_{\mathbf{0}} = \begin{cases} [P_{a'b}(\omega)]^T & N = \{1,3,5,\dots\} \\ [P_{ab'}(\omega)] & N = \{2,4,6,\dots\}, \end{cases}$$

(9)

where a' and b' denote the virtual source depth on V^α or V^β, respectively, depending on which array is appropriate, and the superscript $(^T)$ denotes the transpose operation. In this analysis, subscripts appearing with the prime notation $(')$ indicate the virtual source element location. The iterated field is a product of matrices

$$\mathbf{Q}_{\mathbf{N}} \equiv \left[\prod_{n=1}^{N} \mathbf{\Gamma}^{(-\mathbf{1}^{\mathbf{N-n}})} \right],$$

(10)

where $\mathbf{\Gamma}$ represents the appropriate form of the acoustic response

$$\mathbf{\Gamma}^1 \equiv A(\omega) \mathbf{P} \mathbf{\Delta}_{\mathbf{z}}^{\beta}$$

$$\mathbf{\Gamma}^{-1} \equiv A(\omega) \mathbf{P}^{\mathbf{T}} \mathbf{\Delta}_{\mathbf{z}}^{\alpha}.$$

(11)

These expressions are generalized to include a normalization factor, $A(\omega)$, whose meaning is addressed in Sec. IIB2c. Diagonal matrices $\mathbf{\Delta}_{\mathbf{z}}^{\alpha}$ and $\mathbf{\Delta}_{\mathbf{z}}^{\beta}$ are included to account for the discretization of the array depth sampling,

$$\mathbf{\Delta}_{\mathbf{z}}^{\alpha} = \begin{bmatrix} \Delta z_1 & 0 & \cdots & 0 \\ 0 & \Delta z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta z_A \end{bmatrix},$$

(12)
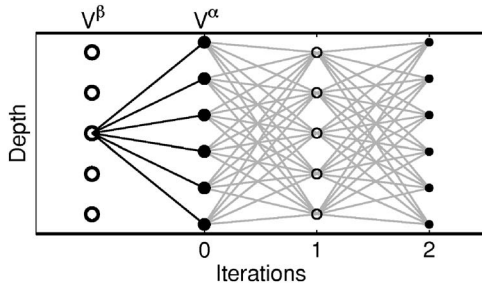
FIG. 2. Schematic of the iteration procedure for two iterations generalized to arrays of differing geometry. Elements of V^α and V^β are indicated by closed and open circles, respectively. Larger circles denote actual array elements while smaller circles mark virtual sample locations. The pressure field sampled by V^α due to a source at a chosen element depth on V^β (indicated by the black lines) is chosen as the starter field. Bear in mind V^α is the origin of the iterated field. The array V^β, from which the starter field originates, is drawn to the left of V^α in order to clarify the procedure. Two iterations of the point-to-point array pressure response (gray lines) are applied to propagate the starter field in range.

$$\boldsymbol{\Delta_z^\beta} = \begin{bmatrix} \Delta z_1 & 0 & \cdots & 0 \\ 0 & \Delta z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta z_B \end{bmatrix}.$$

They represent the interelement depth spacings of arrays V^α and V^β, respectively. In the case where the depth sampling is dense, knowledge of the interelement depth spacings is not essential. When the field is densely sampled, deviations of the depth sampling from uniformity result in manageable deviations from orthogonality between the modes. Incorporation of the interelement depth spacings serves to enhance the performance of the iteration method. If the spacings are not known, $\boldsymbol{\Delta_z^\alpha}$ and $\boldsymbol{\Delta_z^\beta}$ are simply unit matrices.

At this point it is instructive to consider an example. Using Eqs. (9)–(13) to explicitly expand the iteration algorithm for $N=2$ yields

$$\hat{\mathbf{P}}_2 = \mathbf{Q}_2 \mathbf{Q}_0 = [\boldsymbol{\Gamma}^1 \boldsymbol{\Gamma}^{-1}][P_{ab'}(\omega)]$$

$$= A^2(\omega)[\mathbf{P}\boldsymbol{\Delta_z^\beta}\mathbf{P^T}\boldsymbol{\Delta_z^\alpha}][P_{ab'}(\omega)]$$

$$\approx A^2(\omega)e^{-i\pi/2}\left(S(\omega)\sqrt{\frac{1}{4\pi R}}\right)^2 [P_{ab'}(3R,\omega)]. \quad (13)$$

The process is shown schematically in Fig. 2. As with Eqs. (6) and (14), a factor of $1/\sqrt{k_m(\omega)}$ has been neglected for each modal component in writing the third line above. This is justified in the case of multiple iterations due to the presence of modal attenuation in any realistic waveguide. With each iteration the effect of the cylindrical spreading error, through the $1/\sqrt{k_m(\omega)R}$ factor, is counteracted by the effect of the exponential decrease resulting from modal attenuation. The net result is that with each iteration the high-order modes tend to attenuate more strongly than the low-order modes.

c. Bandwidth equalization. Because the source signal is embedded in the sampled pressure field, any structure in the source frequency power spectrum will tend to compound with each iteration. As shown in Eqs. (6) and (14), the N-iterated source spectrum is proportional to $S^N(\omega)$. For ex-

ample, in the case of a Gaussian modulated pulse, repeated iteration leads to a reduction of bandwidth. Assuming the source signal is known, such a loss of bandwidth can be overcome by defining the normalization constant of Eqs. (12) and (14) in accordance with the source signal spectrum,

$$A(\omega) = \frac{\sqrt{4\pi R}}{|S(\omega)|}. \quad (14)$$

Including the term in the numerator, this frequency equalization scheme simultaneously compensates for the range dependence of the cylindrical spreading. Due to the term in the denominator, this equalization scheme may prove problematic in the case where the source signal spectrum approaches zero within the bandwidth. Additionally, repeated equalization is subject to limitations associated with noise.

3. The influence of array geometry

The effects of array geometry on TR focusing in an ocean waveguide have been demonstrated.[6,7] Namely, undersampling in depth, both in terms of the sampling interval and the total aperture, leads to reduced focusing performance. Because array geometry effects are compounded with each iteration, TR imaging applications based on the SRI method are even more sensitive to array geometry than standard TR imaging techniques. This section considers three aspects of array geometry that affect the imaging performance of the SRI method: total array aperture, array depth sampling rate, and array tilt. Each aspect is considered separately in order to isolate its individual effects. The following four cases are consider: (1) full-spanning, densely sampling vertical arrays; (2) partial-spanning, densely sampling vertical arrays; (3) full-spanning, sparsely sampling vertical arrays; and (4) full-spanning, densely sampling tilted arrays.

a. Full-spanning, densely sampling vertical array geometries. It has already been demonstrated theoretically that for a full-spanning, densely sampling array [one for which Eq. (4) holds true], iteration results in an accurate approximation of the true pressure response at an equivalent range, particularly with respect to the phase. This is verified in Fig. 3. As the goal is to perform TR focusing, Fig. 3 shows a simulated comparison of TR focusing between an SRI result and the standard modal model for the case of full-spanning, densely sampling arrays. The top panel shows the spatial component of the time-reversed virtual pressure field resulting from eight iterations of the simulated pressure field between the array pair. Eight iterations of the iteration range, $R=500$ m, corresponds to a virtual source at 4.5 km. The simulation is for a narrow-band source at a depth of 45 m. The lower panel shows the TR focus over the same range achieved using the standard modal model for a true source at the same location. Even after eight iterations, there is good agreement between the SRI result and the TR benchmark. It is noted here that all simulation results presented in this paper are carried out for identical simulated Pekeris waveguide propagation media for which the modal components are calculated using the KRAKEN mode model including modal attenuation.[8] The water depth is 100 m, with bottom parameters of sound

2758   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Walker *et al.*: Acoustic focusing in shallow water
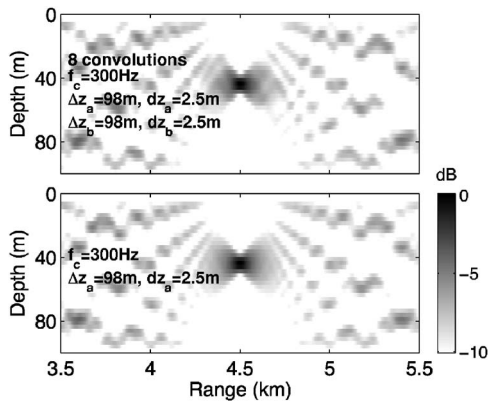
FIG. 3. Comparison of the spatial components of simulated SRI and modal model benchmark time-reversed fields for the case of full-spanning, densely sampling arrays located 500 m apart. The top panel shows the result after eight iterations of the field while the bottom panel is the benchmark result for the modeled field for a source at 4.5 km. The water depth is 100 m and the source depth is 45 m. The depth aperture and element spacing interval of array $V^\alpha(V^\beta)$ are given by $\Delta z_\alpha(\Delta z_\beta)$ and $dz_\alpha(dz_\beta)$, respectively. The source signal carrier frequency is given by f_c.

FIG. 5. Simulated time-domain time-reversed pressure fields corresponding to the array geometries of Fig. 4. All TR results correspond to the TR field measured over a receive array (V^γ of Fig. 1) located at range $r = 4.5$km. The dashed line indicates the expected focal depth. (a) TR result for the benchmark field of Fig. 4(a). (b) TR result for the full-spanning, densely sampled field of Fig. 4(b). (c) TR result for the partial-spanning, densely sampled case of Fig. 4(c). (d) TR result for the full-spanning, sparsely sampled field of Fig. 4(d). (e) TR result for the cleaned full-spanning, sparsely sampled field of Fig. 4(e). (f) TR result for the compensated tilted array case of Fig. 4(f).

speed and density given by $c_b = 1800$ m/s and $\rho_b = 1800$ kg/m³, respectively.

In order to better demonstrate and understand the effects of array geometry, it is useful to consider the time-domain fields. Figure 4 shows a comparison of simulated time-domain fields on arrays of various geometries due to a broadband source at 4.5 km. While panel (a) shows the field from a true source, panels (b)–(f) show iterated fields from virtual sources corresponding to eight iterations of the SRI method. All SRI simulation results are frequency equalized in accor-

FIG. 4. Simulated time-domain pressure fields sampled on arrays of varying geometries from a broadband point source at 4.5 km. All SRI results are for eight iterations of the sampled point-to-point pressure response between arrays separated by 500 m. The source signal is characterized by a Gaussian modulated 200-Hz bandwidth pulse at a carrier frequency of $f_c = 300$ Hz. (a) Benchmark result generated from the modal model for a full-spanning, densely sampling array. (b) SRI result for a pair of full-spanning, densely sampling arrays. (c) Iteration result for the case of partial-spanning, densely sampling arrays. Arrays V^α and V^β spanned 90% and 80% of the water column, respectively. (d) SRI result for full-spanning, sparsely sampling arrays. The sampling interval of the arrays is equivalent to the acoustic wavelength at the carrier frequency in water, $dz = 5$ m. (e) Iteration result for full-spanning, sparsely sampling arrays, cleaned to remove aliased high-order modal components. (f) SRI result for tilted full-spanning, densely sampling arrays compensated to remove the tilt.

dance with the normalization introduced in Eqs. (12) and (15) in order to suppress iterative bandwidth effects.

As a benchmark for comparison, Fig. 4(a) shows the field from a true source, sampled over a full-spanning, densely sampling array. The SRI result for a pair of full-spanning, densely sampling arrays shown in Fig. 4(b) is in good agreement with the true field, particularly with respect to the low-order modal components (the early arrivals). In accordance with the increasing difference in propagating wave number among the higher order modes, the relative weightings of the high-order modal components (late arrivals) are more sensitive to the extra factor of $1/\sqrt{k_m(\omega)}$. Compounding this error over many iterations has resulted in the slight, but observable, difference among the high-order modal components of the iterated field as compared to the benchmark. This effect is moderated by the inclusion of attenuation in the simulation. The time-domain TR results for the SRI full-spanning, densely sampling case, shown in Fig. 5(b), are also in good agreement with the benchmark TR result of Fig. 5(a). To clarify, time reversal is applied by multiplying the phase-conjugated iterated (or benchmark) field by the modal model response from a given range.

b. Partial-spanning, densely sampling vertical array geometries. The case of partial-spanning, densely sampling arrays is addressed in Figs. 4(c) and 5(c). Figure 4(c) shows the time-domain virtually sampled pressure field after eight iterations generated using a 90% spanning array for V^α and an 80% spanning array for V^β. Deviation from the benchmark result [Fig. 4(a)] is apparent over all modal components. This result is expected as partial coverage of the water column results in deviations from orthogonality between all modal components. The resulting TR field in Fig. 5(c) exhibits significant degradation of both the temporal and spatial resolution.

c. Full-spanning, sparsely sampling vertical array geometries. In the case of full-spanning, sparsely sampling ar-

rays, it is possible that orthogonality may exist differentially among the sampled modal components. While the low-order modes may be well sampled in depth, the high-order modes may not. Figure 4(d) shows the compounded effects arising from iterating a sparsely sampled field. The sampling interval of both arrays is equivalent to the wavelength at the carrier frequency, $dz=\lambda_c=5$ m. While there is good agreement among the low-order modal components in comparison to the benchmark, there is an obvious deviation between the high-order modal components. The TR result of Fig. 5(d) displays the characteristic aliasing errors associated with undersampling. One solution for dealing with undersampling errors is to clean the aliased, high-order components from the signal prior to time reversal. Due to the temporal dispersion of the modes, as shown in Fig. 4(e), the high-order components are easily isolated and removed in the time domain. As demonstrated in Fig. 5(e), cleaning the sparsely sampled signal results in a significant improvement in TR focal resolution.

d. Full-spanning, densely sampling tilted array geometries. The final aspect of array geometry considered is array tilt. In the context of this paper, array tilt refers to deviations from vertical aspect among the array elements. In contrast to standard TR, where array tilt has no effect on imaging performance, the SRI method is highly sensitive to tilt. Relative range offsets among the array elements generate phase errors that are amplified by the iteration process. Even a small deviation from verticality of either of the arrays can lead to a drastic degradation of the imaging performance after only a few iterations.

One straightforward way to compensate for array tilt phase errors is to add time delays to the pressure response matrix elements,

$$P_{ab}(R,\omega) \rightarrow P_{ab}(R,\omega)e^{i\omega(\Delta t_a + \Delta t_b)}. \tag{15}$$

In order to do this, the relative range offsets of the array elements must be independently known, either through direct or inferred measurement. Denoting as Δr_a (Δr_b) the range offset of element a (b) with respect to a given reference element, the time delay compensation on element a (b) is given as $\Delta t_a = \Delta r_a / c_s$ ($\Delta t_b = \Delta r_b / c_s$), where c_s is a sound speed characterizing acoustic propagation in the waveguide. Applied in this manner, this form for the time delay can be thought of as an angular beamform operation. Because the group speeds of the low-order modes are more densely packed than those of the higher order modes, choosing a sound speed from among the group speeds of the lowest modes results in the best compensation for the greatest number of modes. Additionally, because they are more highly attenuated, high-order modal contributions tend to be less significant. To a good approximation in a Pekeris waveguide the mode 1 group speed can be estimated by the sound speed of the water column. If the sound speed varies with depth, it is useful to average over depth.

Array tilt simulation results are presented in Figs. 4(f) and 5(f). In this case, array V^α remains vertical while array V^β has been given a characteristic curved tilt profile consistent with a strong current along the ocean bottom. The tilt profile, shown in Fig. 6, has a maximum element range offset
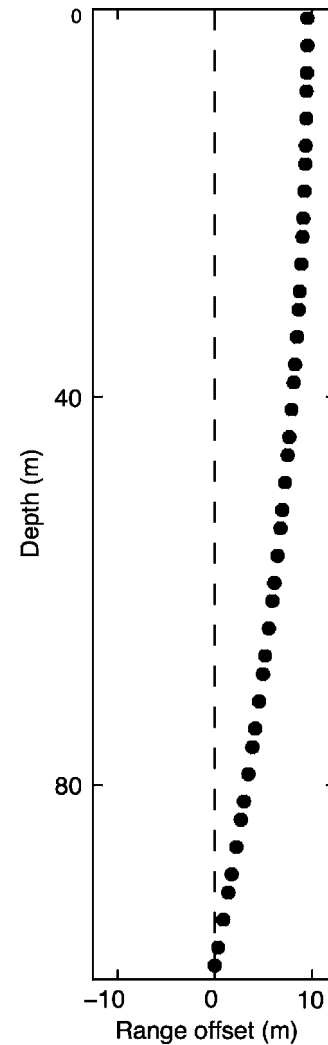
FIG. 6. Array tilt schematic. Dots indicate locations of array elements of V^β relative to the deepest element.

of nearly 10 m (in a 100-m waveguide), or 2 acoustic wavelengths. Iterating the sampled field in this geometry quickly leads to large errors and is not shown. However, applying the tilt compensation method outlined above (assuming a characteristic group speed $c_s=1500$ m/s) to the sampled field under the assumption that the relative array positions (Δr_b) are perfectly known leads to the result shown in Fig. 4(f). As expected, there is good agreement between the low-order modal components (for which the assumed group speed is a good approximation) of the SRI and benchmark fields. The mismatch between the assumed characteristic group speed and the group speeds of the high-order modes leads to small, but observable, accumulated errors among the later arrivals (higher order modal components).

To make the analysis more realistic, the assumption of perfect knowledge of the relative element positions on the tilted array is relaxed. Applying a random uncertainty between −30 and 30 cm to each element range offset (of the tilted array) leads to little observable difference in the SRI time-reversed field (not shown), indicating that small random deviations from perpendicularity are tolerable.
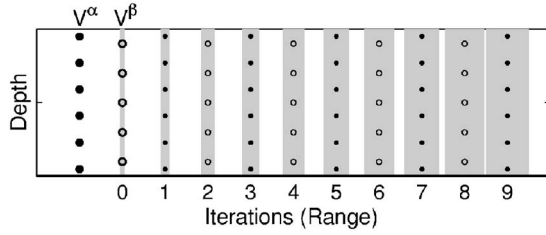
FIG. 7. Schematic of the coverage of the waveguide resulting from combining the SRI method with variable-range TR generalized to the case of arrays of differing geometry. Closed and open circles indicate the element depths of arrays V^α and V^β, respectively. These points represent the ensemble of virtual samples generated by the iteration process. The shaded gray areas indicate the regions of the waveguide to which TR focusing can be achieved when the SRI method is combined with variable-range TR.

4. Combining SRI with variable-range TR

Waveguide invariant theory predicts that interference among modal components leads to a regular pattern of the acoustic intensity as a function of range and frequency.[9–11] One important proven application of invariant theory is range shifting. Reference 4 describes a set of at-sea experiments where the TR focus from a TR mirror was successfully translated in range from the source location by applying a frequency shift to the original received signal on the TR mirror. It was found that applying a frequency shift, $\Delta\omega$, to the frequency domain acoustic field sampled on a vertical array from an acoustic point source at range r_0 leads to a TR focus at the translated range

$$r' \approx r_0\left(1 + \frac{1}{\beta}\frac{\Delta\omega}{\omega}\right), \tag{16}$$

where β denotes the waveguide invariant. Due to the frequency dependence of the depth-dependent amplitudes of the modal components of the acoustic field, the focal resolution of variable-range TR is expected to degrade as the fractional frequency shift, $|\Delta\omega/\omega|$, increases. Empirical evidence suggests that variable-range TR focusing is effective for absolute fractional frequency shifts less than 10%,[4] $|\Delta\omega/\omega| < 0.1$.

By applying variable-range TR to the virtually sampled iterated field, it is possible to extend the TR focusing capability of the SRI technique to include all points of the waveguide. Recall that iteration of the sampled point-to-point acoustic pressure response leads to an ensemble of virtual sources at integer multiples of the iteration range. Placing a conservative restriction on the fractional frequency shift of the carrier frequency, $|\Delta\omega/\omega|_{max}=0.05$, it is possible to generate a high-resolution TR focus at all ranges and depths of the waveguide after eight iterations of the sampled field (assuming $\beta \approx 1$). This can be seen by substituting $|\Delta\omega/\omega|_{max}=0.05$ and $\beta=1$ into Eq. (17) for positive and negative range shifts of ranges corresponding to eight and nine iterations, respectively,

$$9R(1 + 0.05) \approx 10R(0.95).$$

The imaging capability of variable-range TR applied to SRI is schematically presented in Fig. 7. The diagram has been generalized to the case of arrays of differing geometry.
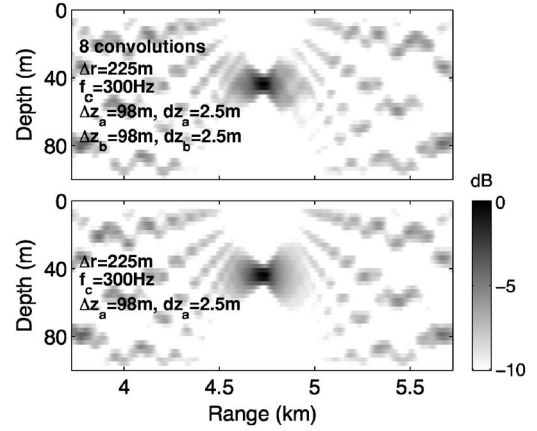
FIG. 8. Comparison of the spatial components of simulated SRI and modal model benchmark range-shifted time-reversed fields for the case of full-spanning, densely sampling arrays. The top panel shows the result for the virtual source at 4.5 km corresponding to eight iterations of the field while the bottom panel is the benchmark result for the modeled field for a source at 4.5 km. In both cases a fractional frequency shift of $\Delta\omega/\omega=+0.05$ has been applied, resulting in the perceived range shift of the fields. The expected range shift, Δr, for a $\beta=1$ waveguide is indicated. The water depth is 100 m and the source depth is 45 m. The depth aperture and element spacing interval of array $V^\alpha(V^\beta)$ are given by $\Delta z_\alpha(\Delta z_\beta)$ and $dz_\alpha(dz_\beta)$, respectively. The source signal carrier frequency (before applying the frequency shift) is indicated by f_c.

Using the iterated virtually sampled field alone, TR foci can be generated at the ensemble of points indicated by the open and closed circles. The closed and open circles correspond to the array element depths of V^α and V^β, respectively. Notice the points span the array depths at multiple integer intervals of the iteration range. Combining SRI with variable-range TR, it is possible to focus at all ranges indicated by the shaded regions at the sampled depths of the corresponding arrays.

Figure 8 shows a simulated variable-range TR result for the case of full-spanning, densely sampling arrays. The top panel shows the spatial component of the time-reversed virtual pressure field resulting from eight iterations. A +5% fractional frequency shift has been applied, resulting in a positive range shift of the TR focus. For the purposes of comparison, the variable-range TR iteration result is accompanied by the variable-range TR result for the standard modal model for the same range and fractional frequency shift (bottom panel). The initial range of the source in both cases (virtual source in the former and real source in the latter) is 4.5 km. The realized focal range shift, $\Delta r \approx 225$ m, is consistent with the expected range shift for a $\beta=1$ waveguide. Figure 9 shows the time-domain range-shifted TR focus (at $r=4.725$ km) corresponding to the simulation of Fig. 8. The resolution of the SRI result [Fig. 9(a)] is comparable to the standard model benchmark result [Fig. 9(b)].

III. EXPERIMENTAL RESULTS

This section presents results of an experimental study of the SRI method. The experiment was conducted under controlled laboratory conditions in a small-scale range-independent waveguide at ultrasonic frequencies. As illustrated in Fig. 1, the point-to-point acoustic response sampled
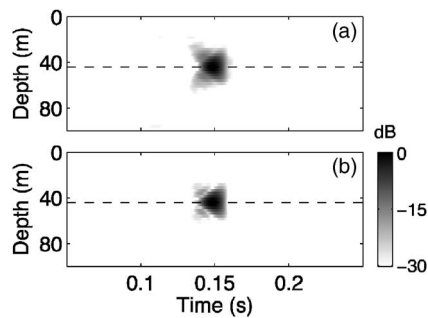
FIG. 9. Simulated time-domain range-shifted time-reversed pressure fields. Each TR result corresponds to the TR field measured over a receive array (V^γ of Fig. 1) located at $r=4.725$ km. (a) Range-shifted SRI result. (b) Range-shifted benchmark result. The time axes indicate the delay with respect to the arrival time of the non-range-shifted TR focus at $r=4.5$ km. The dashed line indicates the expected focal depth.

between a pair of VLAs, V^α and V^β, separated by iteration range R, was iterated to generate a virtual sampling of the point-to-point response at successive integer multiples of the iteration range. The virtually sampled response was then time-reversed with the aim of generating TR foci at the virtually sampled points. A third VLA, V^γ, was positioned down range to measure the time-reversed fields.

A. Setup

The experiment was designed to model the shallow ocean at small scales. The waveguide comprised well-mixed fresh water over a uniform steel bottom maintained at a constant depth over the extent of the waveguide. The working length of the waveguide was roughly 1 m. The water depth was $D=26.5$ mm, resulting in far-field scaling ($R/D \gg 1$). All point-to-point pressure response measurements were sampled using a $\Delta f=1.5$ MHz bandwidth Gaussian modulated source signal at carrier frequency of $f_c=1.5$ MHz. The acoustic wavelength in the water at the carrier frequency was approximately 1 mm, $\lambda_c \approx 1$ mm.

Each VLA, V^α and V^β, consisted of 50 evenly spaced linearly arranged transducer elements spanning 24.5 mm. The element spacing was 0.5 mm ($\approx \lambda_c/2$). Due to the small scale of the setup, it was not possible to determine the array tilt *a priori*. Rather, the tilt was deduced from the performance of the TR results. It was determined that a tilt of 0° for V_α and 0.5° for V_β yielded the optimal TR imaging performance. It is emphasized that the optimization for array tilt is a by-product of the small experimental dimensions. At the larger scales characterizing shallow-ocean applications, the tilt is directly measurable. The iteration range separating the VLAs was $R=82.8$ mm. The VLA used to measure the TR fields, V^γ has identical geometry to V^α.

B. Results

Figure 10 shows the measured SRI TR foci generated from several virtually sampled fields. Panels (a) and (b) are the fields for two different virtual source depths at a range of $r=5R=414$ mm, corresponding to four iterations of the sampled response function. In each case, the dashed lines indicate the expected focal depth. For comparison, the mea-
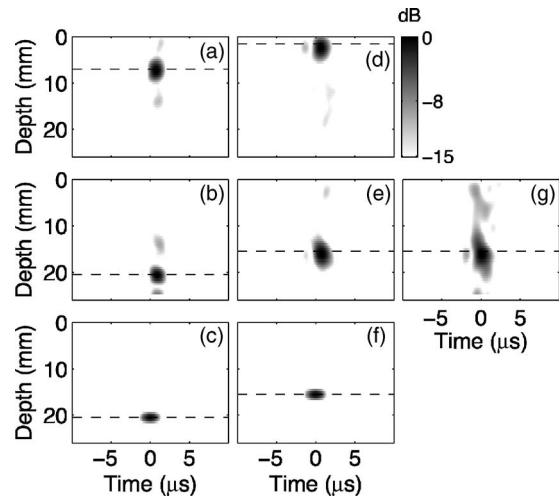
FIG. 10. Experimental time-domain time-reversed pressure fields measured over a receive array (V^γ of Fig. 1) at the indicated ranges. In all cases, the dashed line indicates depth of the source from which the TR field originated (real or virtual). Panels (a) and (b) show the TR foci produced by iterated fields (four iterations) corresponding to virtual sources at $r=414$ mm$=5R$. Panel (c) is the TR result for a real source at the same range. Similarly (d) and (e) show the TR foci produced by iterated fields (nine iterations) corresponding to virtual sources at $r=828$ mm$=10R$ with panel (f) giving the TR result for a real source at the same range. Panel (g) corresponds to panel (e). In this case, array tilt was not accounted for. The source signal bandwidth and carrier frequency were $f_c=1.5$ MHz and $\Delta f=1.5$ MHz, respectively.

sured TR focus due to a real source at the same range, $r=414$ mm, is included in panel (c). Similarly, panels (d) and (e) show SRI TR focus results due to virtual sources at $r=10R=828$ mm (nine iterations). Again, the measured TR focus due to a real source at the same range is presented in panel (f) for comparison. The experimental SRI TR result in panel (g) resulted from application of the iteration procedure without accounting for array tilt. The tilt was relatively small, accounting for a total maximum absolute interelement range displacement of $|\Delta r_b|_{max}=0.2$ mm$\approx \lambda_c/5$. After nine iterations, however, even such a small tilt results in a significant degradation of imaging performance. Noting that the temporal extent of the TR foci remained constant over many iterations, it is encouraging that the frequency equalization scheme of the iteration procedure resulted in negligible iterative bandwidth loss. In experimental applications, it is important to consider the frequency response of the transducer devices. As it is embedded in the sampled point-to-point pressure response, the transducer response will be propagated with each iteration.[12] Ideally, transducer responses could be measured and accounted for in the equalization scheme. Similar experimental results were achieved at all depths spanned by the VLAs out to nine iterations. The usable working range limited the experiment to nine iterations. The results suggest that the virtual field could have been propagated out further.

The SRI method was combined with the variable-range-TR method of Ref. 4 to shift the TR focus from integer multiples of the iteration range to intermediate ranges. As mentioned, it is possible to range shift the TR focus to all intermediate ranges after eight iterations in a $\beta=1$ waveguide assuming a conservative 5% range shift maximum. Figure 11 shows an experimental range-shifted TR focus re-
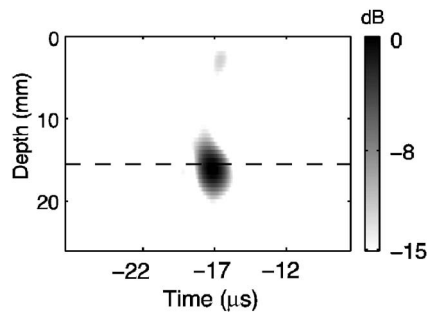
FIG. 11. Experimental time-domain range-shifted TR focus measured with a receive array (V^γ of Fig. 1) at $r = 801$ mm. The time axis indicates the delay with respect to the arrival time of the non-range-shifted TR focus at $r = 828$ mm (nine iterations). The dashed line indicates the expected focal depth.

sult. In this case a fractional frequency shift of $\Delta\omega/\omega = -3.33\%$ ($\Delta f = -4.884 \times 10^4$ Hz) has been applied to the iterated field that produced the TR focus at $r = 10R = 828$ mm (nine iterations) shown in Fig. 10(e). Figure 11 shows the time-domain field at the expected range-shifted focus range, $r = 801$ mm (assuming a $\beta = 1$ waveguide). Similar experimental range-shifted TR focus results were achieved at all sampled intermediate ranges and depths between $r = 744$ mm (eight iterations) and $r = 828$ mm (nine iterations).

IV. SUMMARY AND DISCUSSION

A method for using a pair of vertical transducer arrays to ensonify a spatially compact region at any location of a shallow, range-independent ocean waveguide has been presented. The method termed sampled response iteration (SRI), is fully self-adaptive, requiring no *a priori* knowledge of the environment other than that the medium is range independent and static. Furthermore, in contrast to existing time-reversal (TR) focusing methods, the technique does not require a down-range source.

Denoting the initial sampling range separating the arrays as the iteration range, the sampled point-to-point acoustic pressure response between the arrays has been multiply iterated to generate accurate estimates of the point-to-point acoustic response projected over the arrays at integer multiples of the iteration range. In both simulation and experiment, the procedure resulted in a virtually sampled acoustic field that was time-reversed to produce high-intensity, high-resolution acoustic focusing at any of the virtually sampled locations. The SRI method has been combined with the variable-range-time-reversal technique proposed by Song *et al.*[4] to demonstrate the potential to produce time-reversal foci at ranges other than the virtually sampled ranges. The SRI method has proven effective experimentally under controlled laboratory conditions out to nine iterations.

As the process is iterative, array geometry effects assume great importance. Any deficiencies in the sampling geometry of the arrays compound with each iteration. The method is effective over many iterations for full water column spanning geometries, even in the case of sparse array sampling. In the case where one or both of the arrays are partially spanning, the effectiveness of the iteration procedure is limited to fewer iterations, with the limit depending on the extent of the array coverage. Of particular importance

is the issue of array tilt. For the iteration procedure to be effective, array tilt must be compensated for, requiring that the arrays be equipped with tilt measurement devices. So long as the arrays can be rendered vertical on average, small uncertainties in tilt compensation are tolerable.

Because the iteration invokes range invariance, the analysis presented in this paper has been limited to range-independent applications. However, there is reason to suspect that the method may be extendable to weakly range-dependent environments, suggesting a possible avenue of future study. The SRI method has potential as both a passive and active sonar application. As a passive sonar application, SRI is a natural candidate for matched-field signal processing (MFP) methods. For example, a pair of low-power arrays separated by a short-range interval could be used to self-adaptively generate replica fields at long ranges. Because the arrays are low power, they would remain undetectable by long-range targets. It may be possible to make the method completely passive using the ambient noise field. As demonstrated by Roux *et al.*,[13] cross correlation of the ambient noise acoustic field over the array elements leads to the array pressure response between the elements. The signal from any source, then, could be cross correlated and iterated to determine the source location. As an active sonar application, the demonstrated ability of SRI, when combined with time reversal, to ensonify spatially compact regions of the shallow ocean makes it a promising candidate for directed sonar. Additionally, as it is a method for generating a large virtual range aperture, the SRI method may prove useful as a mode extraction application. Of particular interest is the potential for extracting the lowest order modes requiring the greatest range aperture.

[1]S. N. Wolf, D. K. Cooper, and B. J. Orchard, "Environmentally adaptive signal processing in shallow water," *Oceans'93, Engineering in Harmony with Ocean Proceedings*, **1**, 99–104 (1993).

[2]A. B. Baggeroer, W. A. Kuperman, and P. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **1**(4), 401–424 (1993).

[3]G. A. Grachev, "Theory of acoustic field invariants in layered waveguides," Acoust. Phys. **39**(1), 748756 (1993).

[4]H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," J. Acoust. Soc. Am. **103**, 3224–3240 (1998).

[5]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).

[6]P. Roux, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "A nonreciprocal implementation of time reversal in the ocean," J. Acoust. Soc. Am. **116**, 1009–1015 (2004).

[7]W. A. Kuperman, W. S. Hodgkiss, and H. C. Song, "Phase-conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).

[8]M. B. Porter, "The KRAKEN normal mode program," SACLANT Undersea Research Center, La Spezia, Italy, 1991.

[9]L. M. Brekhovskikh and Yu. Lysanov, *Fundamentals of Ocean Acoustics* (Springer, Berlin, 1991), pp. 139–145.

[10]S. D. Chuprov, "Interference structure of a sound field in a layered ocean," in *Acoustics of the Ocean: Current Status* (in Russian), edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982), pp. 71–91.

[11]G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**(5), 2454–2468 (1999).

[12]W. J. Higley, Private correspondence.

[13]P. Roux, K. Sabra, and W. A. Kuperman, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Walker *et al.*: Acoustic focusing in shallow water   2763

Modeling distortion product otoacoustic emission input/output functions using segmented regression

Bryan Goldman
Fred Hutchinson Cancer Research Center, Seattle, Washington 98109

Lianne Sheppard
Department of Biostatistics, University of Washington, Seattle, Washington 98195 and Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington 98195

Sharon G. Kujawa
Department of Otology and Laryngology, Harvard Medical School and Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114

Noah S. Seixas
Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, Washington 98195

Distortion product otoacoustic emissions (DPOAEs) are low-level acoustic signals, the detection of which involves extraction from a background of noise. Boege and Janssen [J. Acoust. Soc. Am. **111**, 1810–1818 (2002)] described a method for modeling the presence and growth of these responses. While improving growth function parameter estimation, this technique excludes a significant fraction of the data (especially low-level responses), and relies on ad hoc model fit acceptance criteria. The statistical difficulties associated with these limitations are described, and a weighted segmented linear regression model that avoids them is proposed. A simple test is presented for the presence of DPOAE growth. This technique is compared to that of Boege and Janssen in a dataset of 9 556 input/output (I/O) functions collected over 4 years on 866 ears from 379 construction apprentices and 63 age-matched controls. Comparisons are made on the entire dataset and within audiometric hearing loss categories. Segmented regression avoids the statistical pitfalls of the previous method, allows estimation of the threshold and slope of auditory response on a far greater number of I/O functions, and improves estimation of these parameters in this dataset. The potential for this method to yield more sensitive metrics of hearing function and compromise is discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2258871]

I. INTRODUCTION

First reported by Kemp (1978), otoacoustic emissions (OAEs) reflect acoustic energy produced by cochlear outer hair cells, either spontaneously or in response to an external stimulus. The distortion product OAE (DPOAE) has generated considerable interest as a sensitive and frequency-specific indicator of outer hair cell function (Brownell, 1990; Stypulkowski, 1990; Kujawa *et al.*, 1992, 1994; Avan and Bonfils, 2005). DPOAEs are produced by simultaneously presenting two pure tones at frequencies f1 and f2 and intensities L1 and L2 to the ear. In normally functioning human ears, this yields a family of distortion components. Under many stimulus conditions, the most robust of these, and thus the most commonly monitored, corresponds to the frequency 2f1−f2. Stimulus parameters that affect the level, or strength, of the DPOAE include the intensity, the level difference (L1−L2), and the frequency ratio (f2/f1) of the two pure-tone stimuli used to produce it (e.g., Brown *et al.*, 1992; Whitehead *et al.*, 1995a, 1995b; Moulin, 2000).

At the ear canal site of their detection, DPOAEs are present in a background of noise, and may be easily identified as responses only if their levels exceed those of other frequency components present in the surrounding canal spectrum. Since both the emission and the noise floor are measured with error, not all OAE measurements with a positive signal to noise ratio (SNR) represent actual emissions, nor is the reverse true. Rather, the SNR is positively related to the probability that an OAE measurement represents an actual emission (Kimberley *et al.*, 1997). Nonetheless, a SNR criterion approach is commonly used to identify valid DPOAE responses (Vinck *et al.*, 1996; Boege and Janssen, 2002; Gorga *et al.*, 2003; Oswald and Janssen, 2003). In other words, the level of the noise floor is treated like a fixed limit of detection. Since, however, noise floor level varies with frequency, and depends heavily upon the environment, equipment, and protocol used in testing, as well as the physiologic noise created by the individual being tested, the noise floor behaves like a random variable. The statistical implications of this with respect to growth function modeling are explored in detail below. Although measurement-based stopping rules may be used in an effort to reduce the noise floor level (e.g., Gorga *et al.*, 2003), they do not eliminate the effect of variability across subjects and testing conditions.
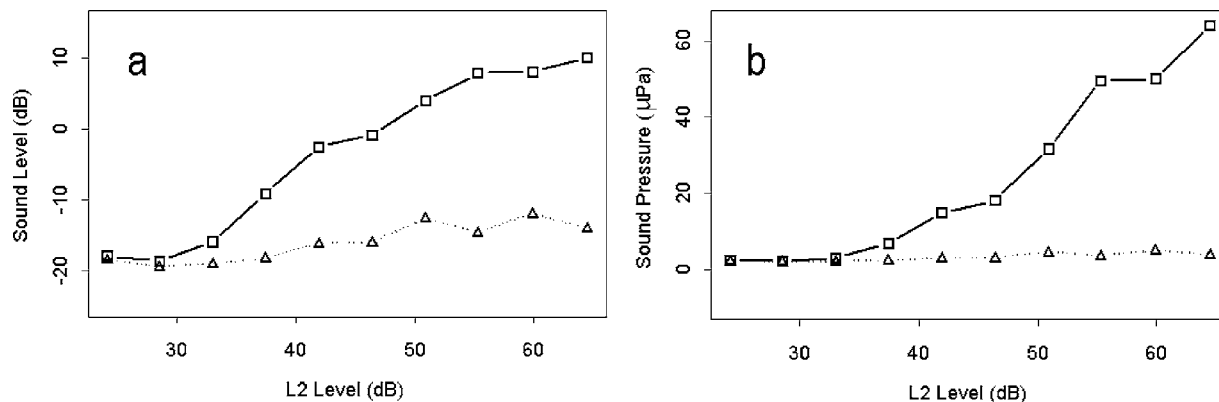
FIG. 1. Sample I/O function (f2/f1 = 1.22, L1−L2 = 10 dB) from a subject with normal hearing, plotted both in dB (a) and in μPa (b). These units are related as follows: $P = 20 \times 10^{L/20}$, where L is the level in dB and P is pressure in μPa. In this example, the DP emission (□) diverges from the noise floor (△) at an L2 level of around 33 dB, indicating the approximate beginning of the growth function.

Such rules also induce an association between DPOAE level and noise floor, raising other statistical concerns discussed below.

Several cross-sectional studies have found correlation between DPOAE level at a single f1 primary stimulus level (usually between 60 and 75 dB) and standard audiometric thresholds, both among normal and hearing-impaired ears (Avan and Bonfils, 1993; Gaskill and Brown, 1993; Kimberley *et al.*, 1994; Gorga *et al.*, 1997; Lind, 1998). Additionally, data collected longitudinally in both animals and humans suggest that declines in DPOAE strength may correlate with future, as well as concurrent hearing loss (Littman *et al.*, 1998; Stavroulaki *et al.*, 2001; Lucertini *et al.*, 2002). Indeed, some reports have described shifts in DPOAEs that precede behavioral threshold shifts in ears with presumed cochlear insult from ototoxic drugs (Littman *et al.*, 1998) and noise (Lonsbury-Martin and McCoy, 1992; Seixas *et al.*, 2005; Sallustio, 1998). Such findings suggest that DPOAEs may provide early warning of future sensory hearing loss. These preclinical changes will likely be first evident for low-level responses (Kujawa *et al.*, 1992, 1997; Mills *et al.*, 1993). As a result, methods which reliably extract low-level responses from their surrounding background of noise should yield the most sensitive metrics of compromise.

The difficulties encountered by current DPOAE extraction methods are depicted in Fig. 1. DPOAE response growth is plotted as a function of increasing stimulus level over a range capturing its fluctuation, eventual emergence from the noise floor, and roughly linear growth. At higher levels, DPOAE growth often becomes nonmonotonic, as cochlear active processes saturate (Kimberley *et al.*, 1997). At lower levels, DPOAE level is dominated by the noise floor, making it unclear to what extent, if any, the measured value represents an emission.

Although several methods for modeling and summarizing the growth function have been developed in the past decade (Nelson and Zhou, 1996; Vinck *et al.*, 1996; Kimberley *et al.*, 1997; Boege and Janssen, 2002; Oswald and Janssen, 2003; Gorga *et al.*, 2003; Neely *et al.*, 2003; Shaffer *et al.*, 2003), much recent work has been based upon the method of Boege and Janssen (2002). They proposed a simple linear regression model to describe the relationship

between L2 and DPOAE pressure. The x intercept of this line is interpreted as an extrapolated DP threshold (L_{edpt}), or the theoretical minimum L2 needed to elicit a cochlear response. This method offers the advantage over fixed SNR-based criterion methods that it is flexible enough to obtain ear-specific point estimates and standard error estimates of growth function parameters. By using the information contained from multiple DPOAE responses, it reduces the impact of DPOAE measurement error on parameter estimation. The parameter of primary interest (L_{edpt}) has shown modest correlation with audiometric thresholds, suggesting that this model may permit estimation of pure-tone thresholds from DPOAE data.

More recently, Oswald and Janssen (2003) refined this method by upweighting points in proportion to their SNR and downweighting points at the upper end of the I/O function, where outer hair cell saturation can cause departures from linearity in the growth function. Under the same set of data inclusion and model acceptance criteria, this weighted regression led to more model fits being accepted, thereby allowing threshold and slope estimation in a greater number of I/O functions. Gorga *et al.* (2003) showed improved model performance with some changes to the model acceptance criteria, and when I/O functions were recorded over a broader intensity range.

All of these variations on the L_{edpt} method of Boege and Janssen involve the exclusion of all DPOAE measurements that fail to achieve a SNR of at least 6 dB. As discussed, the SNR is related to the probability that a measured value at the emission frequency reflects an actual auditory response. One set of results estimated this probability for SNRs of 3 and 6 dB at 0.82 and 0.95, respectively (Kimberley *et al.*, 1997). Although no published studies have validated these estimates over a range of OAE testing conditions, they nonetheless imply that application of this 6 dB threshold involves the exclusion of valid data. Furthermore, points excluded by this SNR-based criterion are preferentially located in the low-level portion of the I/O function, eliminating values that may provide important insights regarding cochlear functional status, and excluding valid and highly influential data points from the analysis of the growth function. In the presence of measurement-based stopping rules, this method encounters

further statistical difficulties in that points are excluded based on the value of a variable (noise floor) that is correlated with the dependent variable (DPOAE level). Such informative missingness is known to yield biased regression parameter estimates (Seber and Lee, 2003). A method that avoids these pitfalls and makes more efficient use of the information contained in low-level responses, and is therefore more sensitive to cochlear compromise, is therefore desired.

The L_{edpt} method also applies criteria that accept as valid only those model fits in which the slope and Pearson correlation coefficient between DPOAE pressure and L2 are sufficiently large, and the standard error of L_{edpt} is sufficiently small. Since the Pearson correlation coefficient is highly sensitive to outliers, it may not be a reliable measure of association when so few points (typically 3–10) are modeled. Furthermore, although this method sets a lower bound for the growth function slope estimate, it does not offer any formal test for the presence of a growth function, which may limit its ability to reliably distinguish those cases in which a true response is present from those in which it is not. When growth function parameter estimation is of primary interest, a modeling technique that explicitly estimates the onset of the growth function and specifically tests for its presence without relying on ad hoc selection criteria would be preferred.

In this paper, we present a segmented regression model for the analysis of I/O function data collected from a large cohort of normal and noise-damaged ears. Segmented regression describes a class of models in which the relationship between the outcome and the predictors changes abruptly at a threshold value of one of the predictors (Muggeo, 2003; Hinkley, 1971), and has many applications in the biomedical literature, especially in describing dose-response relationships (Ipp *et al.*, 1990; Berman *et al.*, 1996; Kuchenhoff and Ulm, 1997; Pastor and Guallar, 1998; Luwel *et al.*, 2001; Monti *et al.*, 1980; Pires *et al.*, 2002; Slate and Turnbull, 2000; Teeter *et al.*, 1985). The model presented here assumes that, at near-threshold levels of the DPOAE, its pressure fluctuates randomly about a mean, whereas for stimulus (L2) levels above this point there is a positive linear relationship between the two. Including this changepoint as a parameter in the model allows for direct estimation of ear-specific growth function parameters without the need for data exclusions. By including all low-level responses in the analysis, this method may be more sensitive to preclinical changes in cochlear function. Just as linear regression reduces the impact of DPOAE measurement error in the L_{edpt} method, segmented regression mitigates the increased error associated with low-level responses by "borrowing information" from the entire I/O function. We discuss differences between parameter estimates derived from a purely model-based approach to those obtained after exclusion of some subset of the data. We describe a simple method for using segmented regression results to test for the presence of a growth function. Finally, we compare the parameter estimates obtained by segmented regression and by the L_{edpt} method of Boege and Janssen (2002), both over the entire dataset and within audiometric subgroups.

TABLE I. Description of the study population. * Normal hearing at baseline is determined by standard audiometry. Hearing is considered normal if the mean of the audiometric thresholds recorded at 2, 3, and 4 kHz is no greater than 20 dB HL.

	Apprentices	Controls	Total
Subjects	379	63	442
% Male	89.4%	54.0%	84.4%
Mean (SD) age	27.3 (6.9)	27.3 (4.1)	27.3 (6.6)
Ears	741	125	866
% Normal Baseline Hearing*	83.1%	96.0%	85.0%

II. METHODS

A. General

As part of a longitudinal study on the effect of occupational noise exposure on various hearing outcomes, audiometric information was collected over 4 years for a group of apprentices in various construction trades and graduate student controls of similar age. All lived in the Seattle area. All subjects underwent otoscopy, tympanometry, pure-tone threshold audiometry, and detailed DPOAE evaluation. Tests were conducted by CAOHC-certified audiometric technicians and audiologists; all results were reviewed by an audiologist (see Seixas *et al.*, 2004, 2005 for details).

B. DPOAE dataset

A total of 10 589 I/O functions at f2=2, 3, 4, and 6 kHz were collected over 4 years on 907 ears in 454 subjects. DPOAEs were measured using a commercially-available system (Bio-Logic Scout® AuDX) in a quiet room (levels generally less than 70 dBA), verified by periodic sound-level monitoring. Each I/O function was comprised of 10 points, corresponding to target L1 intensities of 35−80 dB in 5 dB steps, with f2/f1=1.22 and the target L1−L2 difference =10 dB. Noise floors were sampled at ±50 Hz of the 2f1−f2 frequency.

All I/O functions were excluded where testing irregularities led to a change of probe and/or technician during the testing session. Also excluded were all I/O functions recorded on ears with nonintact tympanic membranes or in which tympanometry indicated middle ear pressures more negative than −100 daPa. Individual DP responses were excluded where the achieved L1−L2 difference deviated from the target setting of 10 dB by more than 2 dB. Finally, 16 I/O functions were removed because fewer than 5 points remained after the other exclusions. The final dataset comprised 95 489 DP emissions from 9 556 I/O functions on 866 ears in 442 subjects. The number of study visits for each subject ranged from 1 to 4, with a mean of 2.8. The study population is described briefly in Table I.

For 9 157 of these I/O functions, corresponding audiometric thresholds were also recorded. Pure tone air conduction behavioral threshold testing was conducted in a mobile, acoustically-treated audiometric test van by CAHOC-certified audiology technicians (Washington Audiology, Inc., Seattle, WA). Background noise levels in the test van were monitored throughout each testing session using a Quest Bioacoustics Monitor. The test environment met OSHA re-

2766   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Goldman *et al.*: Otoacoustic emission input/output function modeling

TABLE II. Number (percentage) of I/O functions included for analysis under each modeling technique.

Modeling technique	Frequency				Total
	2 kHz	3 kHz	4 kHz	6 kHz	
L_{edpt}	1348 (56.9%)	1372 (57.2%)	1495 (62.4%)	1272 (53.2%)	5487 (57.4%)
Segmented regression	2002 (84.4%)	2039 (85.0%)	2005 (83.6%)	1970 (82.5%)	8016 (83.9%)
Total in dataset	2371	2399	2397	2389	9556

quirements for audiometric test facilities (OSHA, 1983) during all tests, and was in most cases compliant with the more stringent ANSI standard (S3.1-1991) recommended by the National Hearing Conservation Association for audiometric test facilities (NHCA, 1996). Audiometry was conducted on up to six subjects at a time using a Tremetrics RA300 audiometer with TDH-39 headphones. Thresholds were measured in increments of 5 dB.

C. L_{edpt} modeling

A complete description of this method for modeling the I/O function parameters is available in Boege and Janssen (2002). Briefly, it involves three key steps. First, only those DPOAE measurements with an SNR of at least 6 dB are included in the analysis. Second, a simple linear regression of DPOAE pressure against input signal intensity (L2) is fit to those I/O functions with at least 3 remaining points. Finally, those I/O functions in which a linear growth function is present are identified by the following model acceptance criteria: $r^2 \geq 0.8$, standard error of x intercept < 10 dB, slope $\geq 0.2 \, \mu$Pa/dB. Oswald and Janssen (2003) refined this method by applying a weighting scheme designed to downweight points at the extremes of the I/O function, where deviations from linearity often occur.

A weighting scheme similar to that used by Oswald and Janssen (2003) was adopted for the present analysis. Two considerations went into its development. First, since DPOAE pressure is an exponential function of the original data, a mean-variance relationship is present. To avoid any gross violation of the constant variance assumption of the regression model, an inverse-variance weighting scheme was applied in which the DPOAE pressure at each frequency and intensity is assigned a weight inversely proportional to its variance. A mixed effects model was used to estimate this variance from these longitudinal study data (Goldman, 2005; Seixas *et al.*, 2005). This scheme downweights measurements at the high end of the intensity range, where outer hair cell saturation is most likely to cause the growth function to deviate from linearity. Since the goal of this model is to fit the linear portion of the growth function, this inverse-variance weighting scheme has the additional benefit of downweighting those points for which the linear model may be incorrect.

Secondly, a SNR-based weighting scheme was explored. A logistic weighting function was built with its center at 6 dB, and with a shape such that measurements with a SNR of 0 dB are given half the weight of those with a SNR of 6 dB. This resulted in a function that is approximately linear for SNR values from -10 to $+15$ dB, which covers almost all DPOAE measurements in this dataset. All analyses were conducted with the inverse-variance weights alone and with a scaled product of the inverse-variance and SNR-based weights. Since the results of the two schemes were very similar (Goldman, 2005), only the results using the scaled-product weighting scheme are reported.

D. Segmented regression modeling

The segmented regression model proposed for fitting I/O function data consists of a flat line at low L2 levels, an unknown changepoint, and an increasing line at higher L2 levels, corresponding to the growth function. A major difference from the L_{edpt} method is that no DPOAE measurements are excluded from the analysis. This restricted two-segmented linear regression model may be represented as follows:

TABLE III. Breakdown of the number (percentage) of I/O functions included in the segmented regression analysis under each of the inclusion criteria. For algorithm convergence, percentages are out of the entire dataset. For the other criteria, percentages are out of the number remaining after application of the previous criterion.

Inclusion criterion	Frequency				Total
	2 kHz	3 kHz	4 kHz	6 kHz	
Algorithm convergence					
Singular gradient matrix	172 (7.3%)	172 (7.2%)	167 (7.0%)	140 (5.9%)	651 (6.8%)
Step size reduced below min.	76 (3.2%)	96 (4.0%)	163 (6.8%)	105 (4.4%)	440 (4.6%)
Max. No. iterations exceeded	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.0%)
Success	2122 (89.5%)	2131 (88.8%)	2067 (86.2%)	2144 (89.7%)	8464 (88.6%)
Growth function presence (by significant F test)	2062 (97.2%)	2092 (98.2%)	2044 (98.9%)	2050 (95.6%)	8248 (97.4%)
x intercept estimable (by t test on slope estimate)	2002 (97.1%)	2039 (97.5%)	2005 (98.1%)	1970 (96.1%)	8016 (97.2%)

$$E(Y|X = x) = \alpha + \beta(x - \tau)_+,$$

where

$$(x - \tau)_+ = \begin{cases} 0 & \text{if } x < \tau, \\ x - \tau & \text{if } x \geq \tau, \end{cases}$$

and τ is the change point. This equation may be solved via nonlinear least squares using the SPLUS package **nls**. A generalization of standard linear regression, **nls** fits the model

$$y_i = f(\theta, x_i') + \varepsilon_i,$$

where $f(\cdot)$ is not necessarily linear, the parameter vector θ is typically not of the same dimension as the vector of predictors x_i', and all ε_i are independent $N(0, \sigma^2)$ variables. Parameter estimates for this model are found when the sum of squared residuals (RSS)

$$S(\theta) = \sum (y_i - f(\theta, x_i'))^2$$

is minimized. This is done by setting the partial derivatives with respect to θ

$$\frac{\partial S(\theta)}{\partial \theta} = -2 \sum (y_i - f(\theta, x_i')) \frac{\partial f(\theta, x_i')}{\partial \theta}$$

equal to 0. The resulting estimating equations must be solved by numerical optimization, which is accomplished in **nls** using the Gauss-Newton algorithm to step through the parameter space between iterations. The algorithm iteratively fits the model, declaring convergence when the parameter estimates become stable to within a small multiple of their standard error (Bates and Watts, 1988; Chambers and Hastie, 1998). The estimated covariance matrix of the parameter vector is

$$\hat{V}(\hat{\theta}) = s^2 (F'F)^{-1},$$

where s^2 is the estimated least squares error variance and $F = \partial f(\theta, x_i')/\partial \theta$, sometimes called the gradient matrix. For this study, the parameter vector θ consisted of the y intercept (α), the slope of the growth function (β), and the x intercept ($x_{int} = \tau - \alpha/\beta$). The model was parameterized in terms of x_{int} rather than τ to permit direct comparison with the L_{edpt} method, which does not estimate a change point, or starting point of the growth function. It should also be noted that the scaled-product weighting scheme was applied in both the L_{edpt} and segmented regression analyses.

The numerical optimization algorithm may encounter any of several problems, each of which can prevent conver-

TABLE IV. Comparison of the number of I/O functions included in the analysis using the L_{edpt} and segmented regression methods.

		Segmented regression		
		Included	Excluded	Total
L_{edpt}	Included	5218	269	5487 (57.4%)
	Excluded	2798	1271	4069 (42.6%)
	Total	8016 (83.9%)	1540 (16.1%)	9556

gence at the minimum of the RSS. First, in the case of a true minimum that is either very peaky or that yields $\hat{\tau}$ near a boundary of the range of the data, the algorithm may fail to converge, instead moving in miniscule step sizes around the minimum until the step size determined by the algorithm is reduced below the minimum allowed and an error is returned. A less frequent problem occurs when the iterations proceed slowly towards the true minimum, but fail to converge within some predetermined maximum number of iterations. Finally, if the model is inappropriate for the data (or sometimes if initial parameter values are poorly chosen), the gradient matrix will become singular after some number of iterations, leaving the estimating equations without a unique solution. This last error may occur in the presence of extremely noisy data, or when $\hat{\tau}$ converges towards a value outside the range of the data. In this paper, segmented regression model fits are called successful only if no errors were encountered.

The procedure for testing for the presence of a growth function and the validity of inference on its parameters consisted of two parts. First, the absence of a growth function indicates nonresponsiveness of the ear over the range of stimulus intensities presented to it. DPOAE measurements from such an I/O function are therefore said to represent background noise, and are thus adequately summarized as variability around a mean (Lonsbury-Martin and Martin, 1990; Kimberley *et al.*, 1997). This suggests that the weighted F statistic

$$F_0 = \frac{\sum w_i(\hat{y}_i - \bar{y})^2/2}{\sum w_i(\hat{y}_i - y_i)^2/(n-3)},$$

where $\hat{y}_i$ is the fitted value from the segmented regression model and w_i is the weight assigned to the ith DPOAE measurement, may be used to test for the presence of a growth function (Seber and Lee, 2003). Significant evidence of
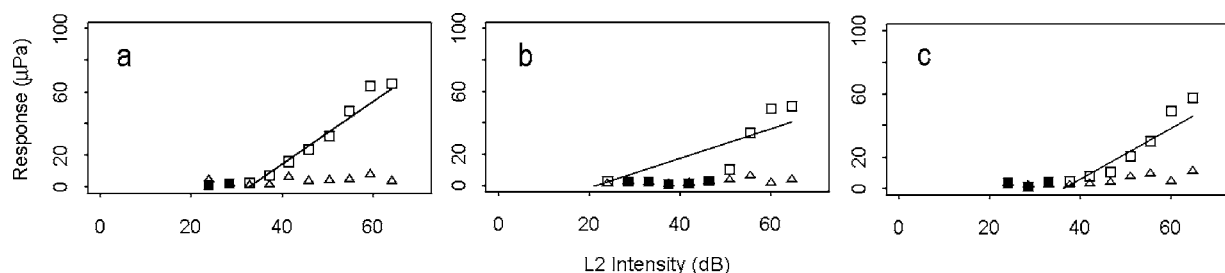
FIG. 2. Examples of successful L_{edpt} model fits, showing DP emissions included ($\square$) and excluded ($\blacksquare$) from the analysis.
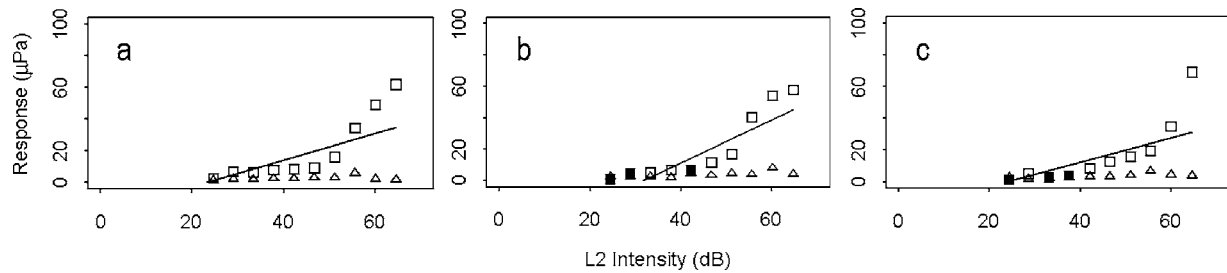
FIG. 3. Examples of L_{edpt} model fits that failed to meet the inclusion criteria, showing DP emissions included (□) and excluded (■) from the analysis.

the presence of a growth function (i.e., that the data are better fit by segmented regression than by a horizontal line at the mean) was declared if the null hypothesis in this F test was rejected at $\alpha < 0.05$.

The second part of the testing procedure addresses the difficulties of performing inference on the x intercept from a linear regression, which is the DPOAE threshold in this analysis. The simple linear regression equation $E[Y|X=x] = \alpha + \beta x$ gives an x intercept estimate of $-\hat{\alpha}/\hat{\beta}$, for which the delta method is often used to estimate the variance as follows:

$$\text{var}\left(\frac{-\hat{\alpha}}{\hat{\beta}}\right) \approx \frac{E[\hat{\alpha}]^2}{E[\hat{\beta}]^2}\left[\frac{\text{var}(\hat{\alpha})}{E[\hat{a}]^2} + \frac{\text{var}(\hat{\alpha})}{E[\hat{\beta}]^2} - \frac{2\,\text{cov}(\hat{\alpha},\hat{\beta})}{E[\hat{\alpha}]E[\hat{\beta}]}\right]$$

(Cox, 1990). This variance estimate is invalid, however, when $\beta = 0$, since in this case the true confidence interval for $-\alpha/\beta$ occupies either the whole real line or the outside of a standard interval (Fieller, 1954; Scheffe, 1970; Young et al., 1997). Inference on the x intercept is valid only when H_0: $\beta = 0$ is rejected; in other words, when the t statistic $|\hat{\beta}/s\hat{e}(\hat{\beta})|$ exceeds some critical value $t_{df,\alpha}$. Accordingly, the x intercept was declared to be estimable only if $\hat{\beta}/s\hat{e}(\hat{\beta}) > 1.90$ (corresponding to the one-sided critical value $t_{7,0.05}$). Note that this t test is based on the segmented regression parameter estimates, and is therefore only meaningful once the F test described above has been rejected.

III. RESULTS

A. Assessment and comparison of model fits

DPOAE thresholds could be estimated from 5 487 I/O functions under the L_{edpt} model and selection criteria, whereas 8 016 I/O functions were included for analysis under the segmented regression method. There was somewhat

more variation across frequencies in the number of I/O functions included under the L_{edpt} model than under segmented regression (see Table II for details).

A breakdown of the number of I/O functions excluded by each inclusion criterion in the segmented regression analysis is shown in Table III. For none of the criteria was there substantial variation across frequencies in the percentage excluded. Overall, the **nls** algorithm encountered an error in 11.4% of I/O functions. Of those remaining, 2.6% did not show significant evidence of a growth function. The x intercept inference criterion rejected 2.8% of those which met the other inclusion criteria.

Table IV compares the inclusion of I/O functions in the analysis by the two modeling techniques. Overall, the two techniques were in agreement regarding inclusion or exclusion of 67.9% of the I/O functions in this dataset. Among those included in the segmented regression analysis, 34.9% were excluded from the L_{edpt} analysis. Of those included in the L_{edpt} analysis, only 4.9% were excluded by segmented regression.

Examples of L_{edpt} model fits that met the inclusion criteria are shown in Fig. 2, and examples of those excluded by these criteria are shown in Fig. 3. Hollow squares depict points included in the analysis, whereas filled in squares show those excluded. Almost all I/O functions that were successfully fit showed clear visual evidence of a growth function. In some cases, however, points that do not appear to be part of the growth function were fit by the line, leading to a smaller slope estimate than the apparent relationship [e.g., Figs. 2(b) and 2(c)]. Many I/O functions excluded by the L_{edpt} criteria also appeared to have growth functions, as in all of the examples in Fig. 3.

Successful segmented regression model fits that met all inclusion criteria are shown in Fig. 4. In general, the model fit appeared to be quite appropriate for these I/O functions. Deviations from linearity, however, were noted at the highest
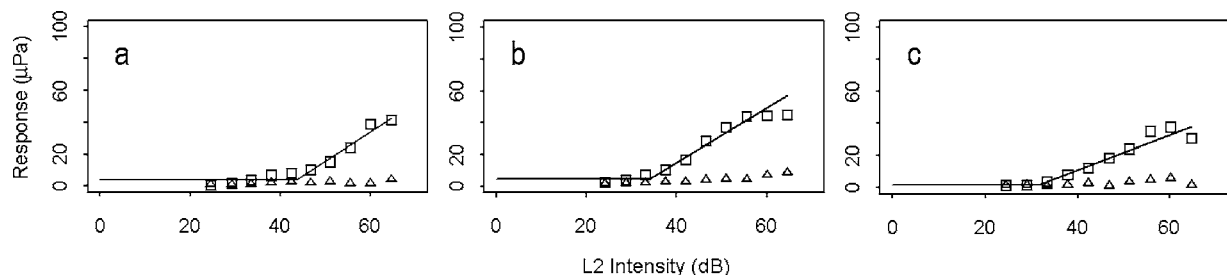
FIG. 4. Examples of successful segmented regression model fits.
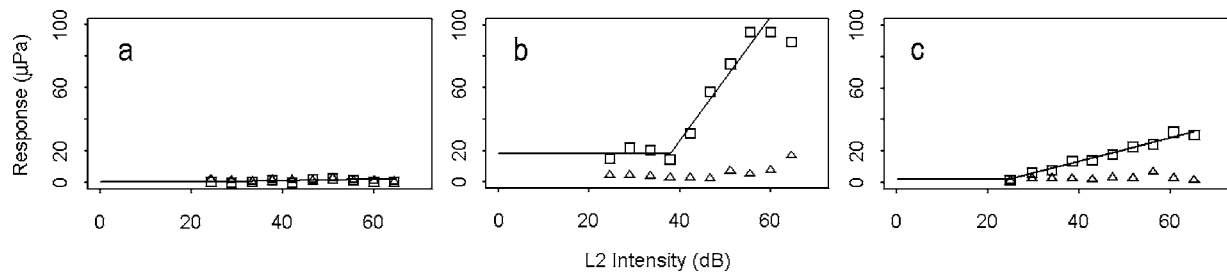
FIG. 5. Examples of segmented regression model fits in which the step size was reduced below the minimum allowed, thereby producing an error.

L2 intensities [e.g., Figs. 4(b) and 4(c)], possibly due to outer hair cell saturation. In another deviation from the model assumptions, some functions showed evidence of a slightly positive slope in the nongrowth region [e.g., Fig. 4(a)].

Examples of functions for which the step size was reduced below the minimum allowed are shown in Fig. 5. Figure 5(c) shows growth with no inflection point, suggesting that this entire I/O function is a growth function, and therefore not appropriately modeled by segmented regression. Other functions, such as in Fig. 5(a), showed so little evidence of growth that the algorithm was unable to declare convergence at any one estimate of the inflection point. This error was also encountered in a few functions for which segmented regression appears to be a reasonable model, but in which a decline in DPOAE pressure was observed at the highest intensities [e.g., Fig. 5(b)]. The functions for which a singular gradient matrix was encountered, shown in Fig. 6, tend to exhibit either growth throughout [e.g., Fig. 6(b)] or no growth whatsoever [e.g., Fig. 6(c)]. A small number show some other irregular pattern [e.g., Fig. 6(a)].

Model fits in which the algorithm successfully converged, but which failed the F test for significant deviation from a simple mean model almost always occurred in I/O functions that have some large deviation from a horizontal line early in the nongrowth portion of the I/O function. Since the inverse-variance weighting scheme upweights points at the lowest intensities, such deviations contributed disproportionately to the sum of squared residuals. Model fits that met the convergence and F test inclusion criteria but not the x intercept inference criterion occurred in I/O functions that showed either very little growth ($\hat{\beta}$ very close to 0), or substantial deviation from the model assumptions of a flat line in the nongrowth portion and of linear growth thereafter [$s\hat{e}(\hat{\beta})$ very large].

A residual plot of the successful L_{edpt} model fits is shown in Fig. 7(a). The standardized residuals plotted on the y axis are the scaled difference between measured DPOAE pressure and that predicted by the L_{edpt} model. Black points were included in the L_{edpt} analysis, whereas gray points were excluded. While most of the excluded points are clustered below $\hat{x}_{int}$, there are many spread out in the region above $\hat{x}_{int}$, all of which have large negative residuals. Residuals at the low end of the I/O function are too large to be shown on this plot. This is because the L_{edpt} model, were it extrapolated below $\hat{x}_{int}$, would predict a large negative DPOAE pressure in this part of the I/O function, which is obviously impossible. The thick solid line represents a local smooth function fit to the residuals from all points included in the L_{edpt} analysis, and the thin solid lines represent the 95% confidence interval of this fit. Since any DPOAE pressure measurement must be positive, all residuals are necessarily positive at $\hat{x}_{int}$ (denoted by the vertical dashed line). Residuals decrease rapidly in the first 5 dB above $\hat{x}_{int}$, and beyond that point a sigmoid residual structure is apparent, suggesting that the linear fit may not completely describe the functional form. The widening of the confidence bands at the high end of the I/O function also suggests, however, that the residual structure in this region is less certain.

Figure 7(b) shows a residual plot of the segmented regression model fits that met all inclusion criteria. Below $\hat{x}_{int}$, a residual positive relationship in the data suggests that DP response may not be entirely independent of L2 intensity prior to the growth function. In the vicinity of $\hat{x}_{int}$, the sigmoidal residual structure is far subtler than that seen for the L_{edpt} model fits. The relationship is nearly flat from 5 dB below to 25 dB above $\hat{x}_{int}$, with a residual mean slightly above 0. At $\hat{x}_{int} + 25$ dB, a negative residual relationship appears in the data. This may be due to outer hair cell saturation causing DP response to level off or decrease at the high end of the I/O function. Although similar to that seen in the residuals from the L_{edpt} model fits, this decrease begins earlier and has somewhat narrower confidence intervals, prob-
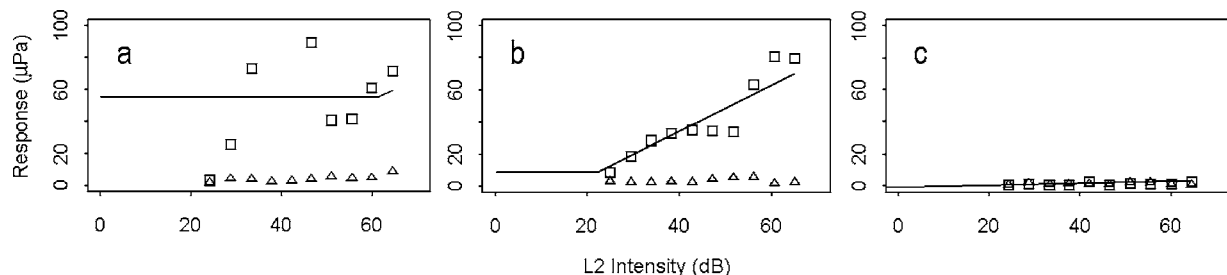
FIG. 6. Examples of segmented regression model fits in which a singular gradient matrix was encountered, thereby producing an error.
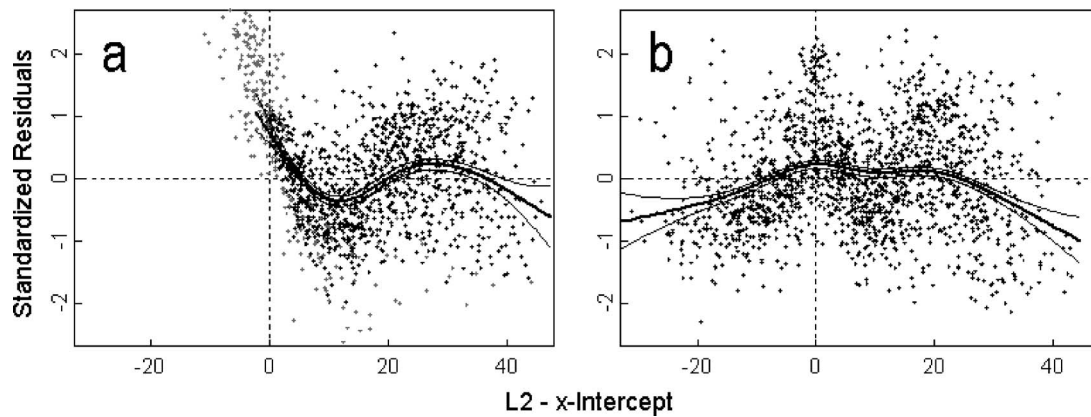
FIG. 7. Standardized residuals for all successful fits using (a) L_{edpt} and (b) segmented regression, plotted against L2$-\hat{x}_{int}$. The solid black lines show a local smooth function of the data banded by its 95% confidence interval. (a) also shows points excluded from the L_{edpt} analysis in gray.

ably because many of the smallest DP responses at the high end of the I/O function were excluded from the L_{edpt} analysis.

B. Comparison of parameter estimates

Summary descriptions of the parameter estimates from the L_{edpt} and segmented regression analyses are given in Table V. In the subset of 5 218 I/O functions which meet the inclusion criteria for both methods, $\hat{x}_{int}$ was 7–8 dB higher at 6 kHz than at the other frequencies, $\hat{\beta}$ was 0.5–0.6 μPa/dB higher at 2 kHz than at 3 or 6 kHz, $s\hat{e}(\hat{x}_{int})$ was lowest at 6 kHz, and $\hat{\beta}$ was lowest at 4 kHz. For both the x intercept and slope, segmented regression appears to yield somewhat larger parameter and standard error estimates than the L_{edpt} method. This also holds true in the comparison of all parameter estimates, although the differences are slightly larger for the point estimates, and much larger for the standard error estimates. These differences do not vary by frequency except that, among slope estimates, they are much smaller at 4 kHz than at the other frequencies. Figure 8 provides a graphical comparison of the parameter estimates obtained under seg-

mented regression and L_{edpt} modeling in the matched subset of 5 218 I/O functions. There is moderately high correlation between them, with a large majority of parameter estimates higher under the segmented regression model.

C. Audiometric subgroups

For 9 157 (95.8%) of the 9 556 I/O functions in this dataset, standard audiometric testing was also performed at the same test visit. Threshold data appeared to be missing at random, as it was equally likely to be missing at each of the four frequencies, and I/O function parameters were similar between those for which audiometric testing was performed and those for which it was not. As shown in Table VI, the percentage of thresholds greater than 20 dB HL in this dataset increased with frequency, especially at 4 and 6 kHz, a frequency region commonly compromised in noise-exposed human ears (NIOSH, 1998; Seixas et al., 2004, 2005). Table VI also shows that, under either method, ears excluded from the analysis were more likely to have elevated behavioral thresholds (exceeding 20 dB HL) than those included. For both methods, the percentage difference between

TABLE V. Mean of growth function parameter point and standard error estimates. The left column summarizes all estimates included in the analysis. The right column summarizes parameter estimates for those 5 218 I/O functions that met the inclusion criteria for both methods.

Method	All estimates			Matched estimates		
	Functions	x Int (SE)	Slope (SE)	Functions	x Int (SE)	Slope (SE)
L_{edpt}						
2 kHz	1348	29.25 (2.04)	1.67 (0.24)	1270	29.77 (1.99)	1.70 (0.25)
3 kHz	1372	30.16 (1.97)	1.18 (0.16)	1284	30.85 (1.93)	1.21 (0.17)
4 kHz	1495	30.47 (1.98)	1.02 (0.13)	1411	31.04 (1.96)	1.03 (0.13)
6 kHz	1272	38.46 (1.77)	1.16 (0.18)	1253	38.60 (1.74)	1.17 (0.18)
Overall	5487	31.95 (1.94)	1.25 (0.18)	5218	32.50 (1.91)	1.27 (0.18)
Segmented regression						
2 kHz	2002	34.59 (2.44)	2.20 (0.37)	1270	33.76 (1.96)	2.16 (0.31)
3 kHz	2039	36.27 (2.61)	1.61 (0.27)	1284	35.52 (2.06)	1.61 (0.23)
4 kHz	2005	35.20 (2.60)	1.20 (0.18)	1411	35.25 (2.03)	1.29 (0.17)
6 kHz	1970	43.27 (2.07)	1.67 (0.29)	1253	43.01 (1.63)	1.57 (0.23)
Overall	8016	37.31 (2.43)	1.67 (0.28)	5218	36.82 (1.92)	1.65 (0.23)
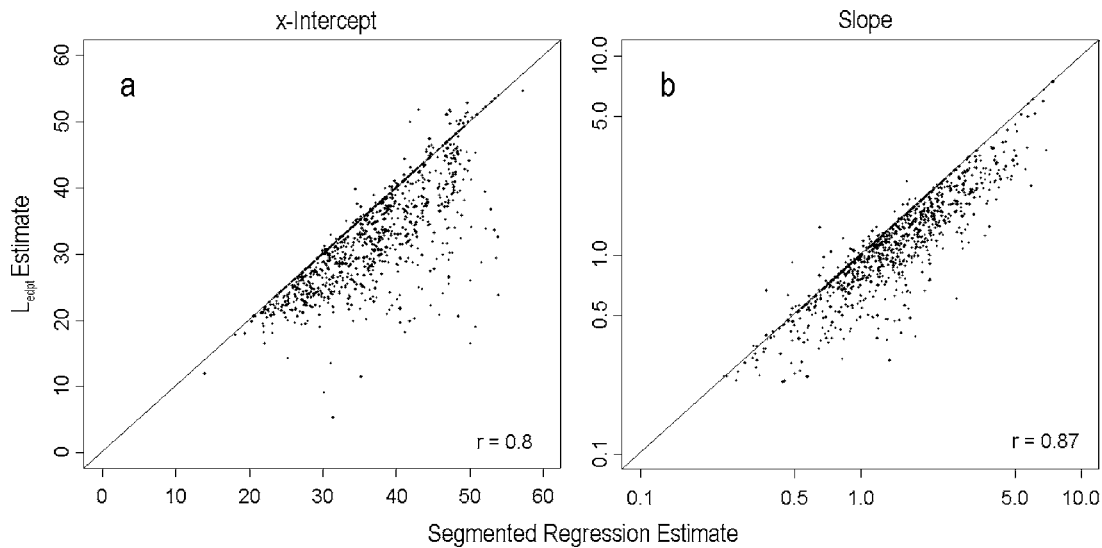
FIG. 8. Comparison of (a) x intercept and (b) slope point estimates by segmented regression and L_{edpt} analysis. The line represents identity. Pearson correlation coefficients are given in the lower right of each plot.

those included in and excluded from the I/O function analysis was greatest at 4 and 6 kHz. Although segmented regression excludes far fewer ears than L_{edpt} overall, a greater fraction of those excluded by segmented regression had elevated behavioral thresholds than of those excluded by the L_{edpt} method. In other words, exclusion by segmented regression was more closely associated with poor hearing than was exclusion by L_{edpt}. This difference was greatest at 6 kHz, where 61.7% of ears excluded under segmented regression had depressed function, as compared to 37.1% of ears excluded from the L_{edpt} analysis.

For 5 012 of the 5 218 DPOAE I/O functions included in both analyses, standard audiometric thresholds were available. Parameter estimates for these I/O functions are broken down by frequency and by audiometric category in Table VII. At 3, 4, and 6 kHz, slope estimates for both methods decrease with increasing thresholds. At all frequencies, x intercept estimates for both methods increase with increasing thresholds, although for both methods at 6 kHz those in the two highest threshold categories are too close to be distinguishable. For both parameters, segmented regression yields larger estimates than the L_{edpt} method at all frequencies and in all audiometric subgroups. Among slope estimates, this difference is substantially greater at all frequencies among ears with audiometric thresholds ≤ 10 dB HL than among those with higher thresholds.

IV. DISCUSSION

Under the segmented regression method presented here, parameter estimates are available for over 46% more I/O functions in the sample dataset than under the L_{edpt} method proposed by Boege and Janssen (2002). This novel approach performs formal statistical hypothesis tests for the presence of a growth function, as well as for whether valid inference can be made on the x intercept of the fitted line. In doing so, it avoids the statistical pitfalls that accompany the use of ad hoc inclusion criteria in the L_{edpt} method. The L_{edpt} method excludes not only a large number of DPOAEs which are highly likely to be valid from the analysis of individual I/O functions, but also entire I/O functions from the grouped analysis based on criteria that are not designed to perform reliably across datasets. Since this method does not apply any x intercept inference criterion, some of the standard error estimates of L_{edpt} that are used as acceptance criteria are likely to be invalid (see Methods for summary). In the present analysis, 3.6% of included L_{edpt} fits fail the x intercept inference criterion applied in the segmented regression analysis.

Some I/O functions were included in the segmented regression analysis where $\hat{\beta}$ was only slightly greater than zero. Table V indicates, however, that segmented regression growth function slope estimates were larger for those I/O

TABLE VI. Percentages of ears with audiometric threshold exceeding 20 dB HL among those included and excluded from the L_{edpt} and segmented regression I/O function analyses, as well as among the overall dataset of 9 157 I/O functions for which audiometric data were available.

Frequency	Overall	L_{edpt}		Segmented regression	
		Included	Excluded	Included	Excluded
2 kHz	4.6	3.0	6.8	3.7	9.1
3 kHz	7.5	3.4	13.0	6.1	15.2
4 kHz	13.2	4.9	27.2	9.9	30.0
6 kHz	24.7	14.0	37.1	16.9	61.4

TABLE VII. Mean of growth function parameter point and standard error estimates for those 5 012 I/O functions that met the inclusion criteria for both methods and had audiometric data available, broken down by categories of audiometric thresholds (in dB HL).

Freq	Pure tone threshold	I/O functions	L_{edpt}		Segmented regression	
			x Int (SE)	Slope (SE)	x Int (SE)	Slope (SE)
2 kHz	≤10	1039	28.96 (1.98)	1.73 (0.24)	33.09 (2.01)	2.23 (0.32)
	>10–20	146	33.09 (2.15)	1.50 (0.22)	36.44 (1.79)	1.78 (0.24)
	>20	37	38.75 (2.03)	2.06 (0.57)	42.23 (1.59)	2.34 (0.36)
3 kHz	≤10	991	29.85 (1.87)	1.26 (0.17)	34.73 (2.05)	1.69 (0.24)
	>10–20	200	34.12 (2.13)	1.02 (0.14)	37.76 (2.02)	1.25 (0.18)
	>20	43	36.45 (2.27)	0.87 (0.14)	40.60 (2.21)	1.09 (0.17)
4 kHz	≤10	1053	30.34 (1.87)	1.09 (0.14)	34.57 (1.99)	1.39 (0.18)
	>10–20	234	33.34 (2.23)	0.85 (0.11)	36.69 (1.98)	0.98 (0.12)
	>20	69	34.85 (2.47)	0.67 (0.12)	40.01 (2.57)	0.81 (0.12)
6 kHz	≤10	576	37.72 (1.63)	1.28 (0.19)	42.37 (1.55)	1.76 (0.25)
	>10–20	463	39.24 (1.86)	1.09 (0.18)	43.61 (1.75)	1.44 (0.22)
	>20	161	39.58 (1.79)	1.02 (0.15)	43.27 (1.68)	1.26 (0.19)

functions excluded from the L_{edpt} analysis than for those included. This suggests that the inclusion criteria for segmented regression were not simply less restrictive than those for the L_{edpt} method, but that the L_{edpt} method fails to identify a large percentage of growth functions in this dataset. Furthermore, within the set of I/O functions included in the L_{edpt} analysis, slope estimates were much smaller among those excluded by segmented regression than among those included. In other words, weaker DP responses are seen among the 269 I/O functions included only by the L_{edpt} method, whereas stronger DP responses are seen among the 2 798 I/O functions included only by segmented regression. It appears, then, that segmented regression more properly distinguishes those I/O functions that show significant evidence of auditory response from those that do not.

In the subset analysis of 5 218 I/O functions included in both analyses, segmented regression gives larger growth function parameter estimates and standard errors than the L_{edpt} method. Since there is no gold standard for defining and estimating the x intercept for DPOAE I/O functions, and since the DPOAE threshold cannot be directly measured (a background sound pressure level of 0 μPa is impossible to achieve), it is difficult to assess which model produces more appropriate parameter estimates. Since the two estimation methods are not typically based on the same set of data points, formal goodness-of-fit tests between these methods are not available. Nonetheless, the fit of each of these models in this dataset may be assessed in the residual plots in Fig. 7. While the residual structure evident in both suggests that neither model has the functional form of the data specified entirely correctly, it is far less pronounced for segmented regression, especially in the immediate vicinity of $\hat{x}_{int}$. For L_{edpt} model fits, the strongest evidence of a residual relationship in the data is seen in this area. In other words, the model breaks down most dramatically in the area most sensitive to cochlear compromise.

The residual plots call into question some of the segmented regression modeling assumptions. In the nongrowth region (below $\hat{x}_{int}$), a positive linear relationship remains in the data. One possible explanation for this lies in the sampling algorithm used by the DPOAE system. The measurement-based stopping rules used by this algorithm induced a relationship between the measured levels of the DP response and of the noise floor. The noise floor may increase slightly as L2 increases, since one component of the noise floor is the noise made by the testing equipment itself. Therefore, a modest positive slope in an I/O function may result primarily from the equipment settings, rather than from any actual auditory response. The residual structure below $\hat{x}_{int}$ would also be present if the true relationship between L2 and DP pressure were a smooth curve rather than the broken stick modeled here, as is likely. As a result of these issues, in testing for the presence of a growth function, the null hypothesis of a flat I/O function appears to be too strong, and may have led to incorrect inclusion of some I/O functions in this analysis.

The residual plots from both sets of model fits show a negative relationship remaining beyond about $\hat{x}_{int}+25$ dB. The residual structure in the data above $\hat{x}_{int}$ suggests that the growth function is not so much linear as sigmoidal, consistent with the saturating nonlinearity of the normal cochlear response described previously by others (see Robles and Ruggero, 2001 for review). The possibility of fitting a logistic function was briefly explored, but there were inadequate data at the high end of each I/O function with which to estimate either the L2 intensity at saturation onset or the upper limit of the DP response.

Despite this evidence of some segmented regression model misspecification, the residual plots shows a much better fit to the data (smaller deviations from the horizontal line) for this model than for the L_{edpt} model. This is especially true over the range $[\hat{x}_{int}, \hat{x}_{int}+25]$, which covers the majority of the growth function, including its onset. It therefore appears better suited to estimate growth function parameters.

Among I/O functions successfully fit by both methods, segmented regression assigned a median of 6 DP emissions to the growth function, whereas the L_{edpt} model was fit to a
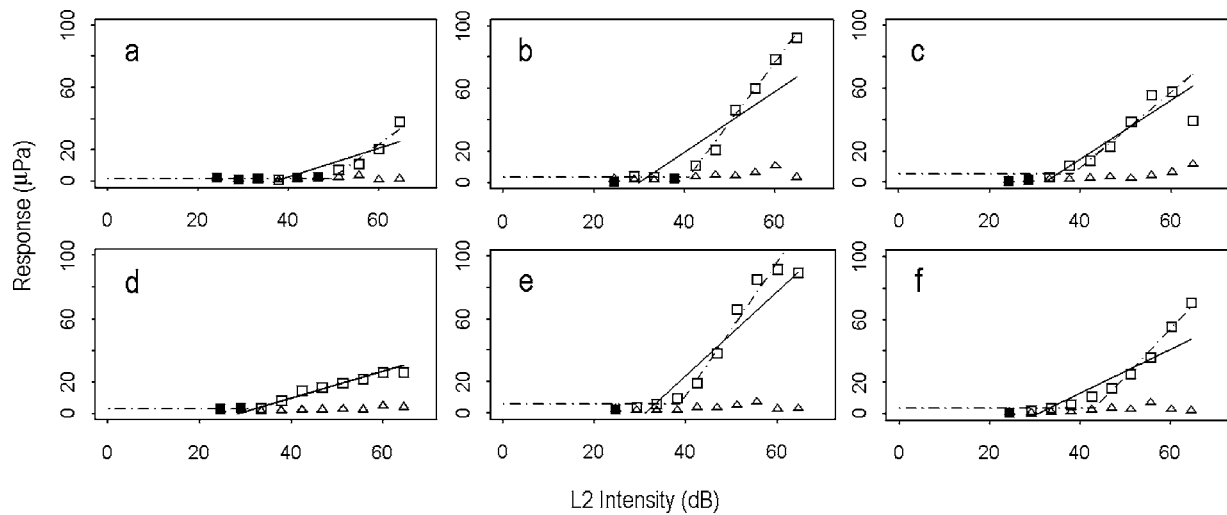
FIG. 9. Comparison of sample L_{edpt} (—) and segmented regression (—··—) fits to DPOAE data. Emissions included (□) and excluded (■) from the L_{edpt} analysis are shown, as well as the noise floor (△).

median of 7 DP emissions. This difference arises from the fact that the L_{edpt} model incorrectly assumes that all points that meet the SNR criterion are part of the growth function and therefore lie on the same line. The effect of this assumption is illustrated in the random sample of I/O functions successfully fit by both models that is plotted in Fig. 9. In Figs. 9(a), 9(b), and 9(f), it is especially apparent that the L_{edpt} method does not properly distinguish between the nongrowth and growth regions of the I/O function, leading to underestimates of both growth function parameters.

Although the segmented regression model presented here provides improved growth function estimation in this dataset, it is unclear how generalizable this result is to other DPOAE datasets. The strength of a DPOAE, and hence the shape of an I/O function, depends on the paradigm used to generate it (Whitehead *et al.*, 1995a, 1995b; see Kimberley *et al.*, 1997 for review). Whereas all published applications of the L_{edpt} method have used DPOAE data generated by a scissor paradigm, the paradigm used to generate this dataset specified a fixed L1−L2 difference of 10 dB. Nonetheless, the comparison of segmented regression and L_{edpt} model fits presented here is applicable to all I/O functions with the features (nongrowth region at low L2 levels, linear growth on the pressure scale at high L2 levels) described by Boege and Janssen (2002).

Whether or not this particular method is validated in other DPOAE datasets, future work in this area should be built upon formal statistical procedures of the sort employed here. Key components of such a procedure include the use of all points in an I/O function and the capability of testing for the presence of a growth function. To the extent that low-level responses contain information about cochlear compromise, the analysis of any set of DPOAE responses benefits from exploitation of this information.

The segmented regression model failed to converge for 11.4% of I/O functions in this dataset, although some of these clearly contained a growth function. In some of these cases, the growth function appeared to comprise the entire I/O function, meaning that data were not collected at suffi-

ciently low L2 levels to observe growth function onset. A simple linear regression may be fit in this case, and tested against H_0: $\beta = 0$ for significant evidence of growth. If, however, the nongrowth portion of the curve were allowed to have positive slope, as proposed above, it is unclear how one would choose H_0 in a test for the presence of a growth function. A preferable direction for future work might involve the collection of data at lower L2 levels to ensure that the nongrowth region has been captured, fitting both simple linear regression and segmented regression to the I/O function, then applying a likelihood ratio test for the changepoint. This procedure would explicitly test for the presence of a growth function without allowing positive slope in the nongrowth region, and its properties have been explored at length in the literature (Feder, 1975; Gallant, 1975; Kim and Cai, 1993; Zhan *et al.*, 1996).

Segmented regression also encountered errors in cases where there was clear growth in DP response followed by a sharp decline at the highest L2 levels. Such extreme nonlinearities clearly require a more flexible I/O function model. Although we explored the possibility of extending this model to allow for a nongrowth region at the highest L2 levels, there were insufficient data in this region to fit such a model. Future work to extend this model to one that estimates the beginning and end of the growth function and tests for the presence of both change points would be desirable, although this will require a richer set of DPOAEs collected over a wider range of L2 levels.

It is also worth noting that improvements in DPOAE measurement techniques may reduce these modeling difficulties. DPOAEs are known to consist of a mixture of reflection and distortion sources, the interaction of which increases the variability in these measurements (Shera and Guinan, 1999; Kalluri and Shera, 2001; Shera, 2004). Shera (2004) has suggested that unmixing these components may substantially reduce the variability in DPOAE measurements. To the extent that this variability currently leads to large deviations from linearity in the growth function, such an approach may im-

Goldman *et al.*: Otoacoustic emission input/output function modeling

prove the ability of segmented regression to detect DPOAE growth, and the sensitivity of the growth function parameter estimates to cochlear changes.

Although previous work with the L_{edpt} method demonstrated fair correlation between extrapolated DPOAE and pure-tone thresholds, similar results would not necessarily be expected in this analysis, since our DPOAEs were not elicited under the scissor paradigm. Also, since audiometric thresholds were measured in 5 dB increments in this study, the analysis of their association with DPOAE thresholds was limited to summaries of the I/O function analyses by audiometric categories. This analysis suggests important differences between the two sets of parameter estimates.

First, at 6 kHz, ears excluded from the segmented regression analysis were far more likely to have depressed function than those excluded under the L_{edpt} model. In this dataset partially comprised of noise-exposed ears, many ears with elevated audiometric thresholds at 6 kHz also showed little or no evidence of a 6 kHz growth function. Although both methods exclude I/O functions from many audiometrically normal ears, these findings are consistent with the hypothesis that exclusions are more likely to be due to the absence of a growth function in segmented regression than in the L_{edpt} analysis. It would nonetheless be desirable to refine this segmented regression procedure such that all I/O functions may be fit and properly tested for the presence of a growth function.

Second, although average parameter estimates from segmented regression exceed those from L_{edpt} in all subgroups in Table VII, this difference is greatest among those ears with audiometric thresholds ≤ 10 dB HL. Segmented regression parameter estimates vary somewhat more across audiometric categories than do those obtained by the L_{edpt} method. This suggests that segmented regression parameter estimates may be more highly correlated with audiometric threshold, although this study was not specifically designed to address this question. One possible explanation for this observation is that subjects with the best hearing may be especially likely to have a DP response with a SNR ≥ 6 dB in the nongrowth portion of the I/O function, in which case L_{edpt} would be especially prone to I/O function parameter underestimation in these subjects.

It would be of great interest if the I/O function parameters, either individually or in combination, were found to reliably predict hearing function and/or future hearing loss. Future work is needed to examine this question more closely, especially in a dataset using the scissor paradigm, as well as to compare the parameters obtained from the L_{edpt} and segmented regression models, in terms of both how well they correlate with other hearing-related outcomes and how well they generalize across levels of hearing function.

V. CONCLUSION

Segmented linear regression represents an improvement over the L_{edpt} method in that it avoids the pitfalls of data-driven exclusions and outcome-based inclusion criteria that are prone to incorrect inference, and makes better use of the information contained in low-level responses. Based on a comparison of model fits, in this dataset of 9 556 I/O functions it appears to provide a more accurate and more complete description of the relationship between DP pressure and L2 intensity, and appears to be more reliable in determining the growth function. Segmented regression also has the advantage of being able to estimate the point of growth function onset, which may be useful as a hearing threshold metric that lies within the range of the data, rather than the extrapolated thresholds examined in this study. Evidence of both model misspecification and misclassification of I/O functions according to the presence of a growth function suggests that more work is needed. It would be desirable to explore whether threshold estimation may be substantially improved by increasing the density and range of intensities in the I/O function, as this would provide not only a richer dataset, but also the possibility of developing models that address nonlinearities in the growth function. Finally, although there is some suggestion that segmented regression parameter estimates covary with audiometric thresholds more than do those obtained from the L_{edpt} model, more work is clearly needed to compare these methods in terms of how well the parameter estimates of each correlate with standard audiometric measures.

Avan, P., and Bonfils, P. (**1993**). "Frequency specificity of human distortion product otoacoustic emissions," Audiology **32**, 12–26.

Avan, P., and Bonfils, P. (**2005**). "Distortion-product otoacoustic emission spectra and high-resolution audiometry in noise-induced hearing loss," Hear. Res. **209**, 68–75.

Bates, D. M., and Watts, D. R. (**1988**), *Nonlinear Regression Analysis and its Applications Models*, Wiley, New York.

Berman, N. G., Wong, W. K., Bhasin, S., and Ipp, E. (**1996**). "Applications of segmented regression models for biomedical studies," Am. J. Physiol. **270**, E723–E732.

Bieler, G. S., and Williams, R. L. (**1993**). "Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity," Biometrics **49**, 793–801.

Boege, P., and Janssen, T. (**2002**). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears," J. Acoust. Soc. Am. **111**, 1810–1818.

Brown, A. M., Gaskill, S. A., and Williams, D. M. (**1992**). "Mechanical filtering of sound in the inner ear," Proc. Biol. Sci. **250**, 29–34.

Brownell, W. E. (**1990**). "Outer hair cell electromotility and otoacoustic emissions," Ear Hear. **11**, 82–92.

Chambers, J. M., and Hastie, T. J. (**1998**). *Statistical Models in S* (Wadsworth & Brooks/Cole, Pacific Grove).

Cox, C. (**1990**). "Fieller's theorem, the likelihood and the delta method," Biometrics **46**, 709–718.

Feder, P. I. (**1975**). "The log-likelihood ratio in segmented regression," Ann. Stat. **3**, 84–97.

Fieller, E. C. (**1954**). "Some problems in interval estimation," J. R. Stat. Soc. Ser. B. Methodol. **16**, 175–185.

Gallant, A. R. (**1975**). "Nonlinear Regression," Am. Stat. **29**, 73–81.

Gaskill, S. A., and Brown, A. M. (**1993**). "Comparing the level of the acoustic distortion product 2f1-f2 with behavioural threshold audiograms from normal-hearing and hearing impaired ears," Br. J. Audiol. **27**, 397–407.

Goldman, B. (**2005**). "Segmented regression for threshold estimation in distortion product otoacoustic emission data" M.S. thesis, University of Washington, Seattle, WA.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). "From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss," Ear Hear. **18**, 440–455.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (**2003**). "Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions," J. Acoust. Soc. Am. **113**, 3275–3284.

Hinkley, D. V. (**1971**). "Inference in two-phase regression," J. Am. Stat. Assoc. **66**, 736–743.

Ipp, E., Sinai, Y., Forster, B., Cortez, C., Baroz, B., Nesher, R., and Cerasi, E., (**1990**). "A glucose reduction challenge in the differential diagnosis of fasting hypoglycemia: A two-center study," J. Clin. Endocrinol. Metab. **70**, 711–717.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kim, H. J., and Cai, L. (**1993**). "Robustness of the likelihood ratio test for a change in simple linear regression," J. Am. Stat. Assoc. **88**, 864–871.

Kimberley, B. P., Hernadi, I., Lee, A. M., and Brown, D. K. (**1994**). "Predicting pure-tone thresholds in normal and hearing impaired ears with distortion product emission and age," Ear Hear. **15**, 199–209.

Kimberley, B. P., Brown, D. K., and Allen, J. B. (**1997**). "Distortion product emissions and sensorineural hearing loss," in *Otoacoustic Emissions: Clinical Applications*, edited by C. Martin, S. Robinette, T. J. Glattke, (Thieme, New York), pp. 181–204.

Kuchenhoff, H., and Ulm, K. (**1997**). "Comparison of statistical methods for assessing threshold limiting values in occupational epidemiology," Comput. Stat. Data Anal. **12**, 249–264.

Kujawa, S. G., Fallon, M., and Bobbin, R. P. (**1992**). "Intracochlear salicylate reduces low-intensity acoustic and cochlear microphone distortion products," Hear. Res. **64**, 73–80.

Kujawa, S. G., Glattke, T. J., Fallon, M., and Bobbin, R. P. (**1994**). "A nicotinic-like receptor mediates suppression of distortion product otoacoustic emissions by contralateral sound," Hear. Res. **74**, 122–134.

Kujawa, S. G., and Liberman, M. C. (**1997**). "Conditioning-related protection from acoustic injury: Effects of chronic differentiation and sham surgery," J. Neurophysiol. **78**, 3095–3106.

Lind, O. (**1998**). "A clinical study on the growth of distortion product otoacoustic emissions and hearing loss at 2 kHz in humans," Scand. Audiol. **27**, 207–212.

Littman, T. A., Magruder, A., and Strother, D. R. (**1998**). "Monitoring and predicting ototoxic damage using distortion-product otoacoustic emissions: Pediatric case study," J. Am. Acad. Audiol **9**, 257–262.

Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). "The clinical utility of distortion-product otoacoustic emissions," Ear Hear. **11**, 144–154.

Lonsbury-Martin, B. L., and McCoy, M. J. (**1992**). "Otoacoustic emissions: Future directions for research and clinical applications," Hear. J. **45**, 47–52.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). "On the detection of early cochlear damage by otoacoustic emission analysis," J. Acoust. Soc. Am. **111**, 972–978.

Luwel, K., Beem, A. L., Onghena, P., and Verschaffel, L. (**2001**). "Using segmented linear regression models with unknown change points to analyze strategy shifts in cognitive tasks," Behav. Res. Methods Instrum. Comput. **33**, 470–478.

Mills, D. M., Norton, S. J., and Rubel, E. W. (**1993**). "Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation," J. Acoust. Soc. Am. **94**, 2108–2122.

Monti, K. L., Koch, G. G., and Sawyer, J. (**1980**). "Segmented linear regression models applied to the analysis of data from a cross-sectional growth experiment," Biometrics **22**, 23–30.

Moulin, A. (**2000**). "Influence of primary frequencies ratio on distortion product otoacoustic emissions amplitude. II. Interrelations between multicomponent DPOAEs, tone-burst-evoked OAEs, and spontaneous OAEs," J. Acoust. Soc. Am. **107**, 1471–1486.

Muggeo, V. M. R. (**2003**). "Estimating regression models with unknown break-points," Stat. Med. **22**, 3055–3071.

Neely, S. T., Gorga, M. P., and Dorn, P. A. (**2003**). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **114**, 1499–1507.

Nelson, D. A., and Zhou, J. Z. (**1996**). "Slopes of distortion-product otoacoustic emission growth curves corrected for noise-floor levels," J. Acoust. Soc. Am. **99**, 468–474.

NHCA (**1996**). "A practical guide to mobile hearing testing and selecting a service provider (Guide #3)." National Hearing Conservation Association: Denver, CO.

NIOSH (**1998**). "Criteria for a Recommended Standard: occupational noise exposure." Cincinnati, O.H., U.S. Department of Health and Human Services/Centers for Disease Control, National Institute for Occupational Safety and Health.

OSHA (**1983**). "Occupational Noise Exposure: Hearing Conservation Amendment; Final Rule," in Fed Reg. U.S. Department of Labor, Occupational Safety and Health Administration: Washington, D.C., pp. 9738–9785.

Oswald, J. A., and Janssen, T. (**2003**). "Weighted DPOAE input/output-functions: A tool for automatic assessment of hearing loss in clinical application," Z. Med. Phys. **13**, 93–98.

Pastor, R., and Guallar, E. (**1998**). "Use of two-segmented logistic regression to estimate change-points in epidemiologic studies," Am. J. Epidemiol. **148**, 631–642.

Pires, A. M., Branco, J. A., Picado, A., and Mendonca, E. (**2002**). "Models for the estimation of a 'No Effect Concentration,' " Environmetrics **13**, 15–27.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Sallustio, V., and Portalatini, P. (**1998**). "Auditory dysfunction in occupational noise exposed workers," Scand. Audiol. **27**, 95–110.

Scheffe, H. (**1970**). "Multiple testing versus multiple estimation, improper confidence sets, estimation of direction and ratios," Ann. Math. Stat. **41**, 1–29.

Seber, G. A. F., and Lee, A. J. (**2003**). *Linear Regression Analysis* (Wiley, Hoboken).

Seixas, N. S., Kujawa, S. G., Norton, S., Sheppard, L., Neitzel, R., and Slee, A. (**2004**). "Predictors of hearing threshold levels and distortion product otoacoustic emissions among noise exposed young adults," Occup. Environ. Med. **61**, 899–907.

Seixas, N. S., Goldman, B., Sheppard, L., Neitzel, R., Norton, S., and Kujawa, S. G. (**2005**). "Prospective noise induced changes to hearing among construction industry apprentices," Occup. Environ. Med. **62**, 309–317.

Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**). "Sources and mechanisms of DPOAE generation: Implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.

Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Shera, C. A. (**2004**). "Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions," Ear Hear. **25**, 86–97.

Slate, E. H., and Turnbull, B. W. (**2000**). "Statistical models for longitudinal biomarkers of disease onset," Stat. Med. **19**, 617–637.

Stavroulaki, P., Apostolopoulos, N., Segas, J., Tsakanikos, M., and Adamopoulos, G. (**2001**). "Evoked otoacoustic emissions: An approach for monitoring cisplatin induced ototoxicity in children," Int. J. Pediatr. Otorhinolaryngol. **59**, 47–57.

Stypulkowski, P. H. (**1990**). "Mechanisms of salicylate ototoxicity," Hear. Res. **46**, 113–145.

Teeter, R. A. (**1985**). "The application of linear piecewise regression to basal body temperature data," Biometrics **7**, 759–773.

Vinck, B. M., De Vel, E., Xu, Z. M., and Van Cauwenberge, P. B. (**1996**). "Distortion product otoacoustic emissions: A normative study," Audiology **35**, 231–245.

Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995a**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L1, L2 space," J. Acoust. Soc. Am. **97**, 2359–2377.

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995b**). "Effects of ear-canal standing waves on measurements of distortion-product emissions," J. Acoust. Soc. Am. **98**, 3200–3214.

Young, D. A., Zerbe, G. O., and Hay, W. W. (**1997**). "Fieller's theorem, scheffe simultaneous confidence intervals, and ratios of parameters of linear and nonlinear mixed-effects models," Biometrics **53**, 838–847.

Zhan, M., Dean, C. B., Routledge, R., Gallaugher, P., and Farrell, A. P., Thorarensen, H. (**1996**). "Inference on Segmented Polynomial Models," Biometrics **52**, 321–327.

Goldman *et al.*: Otoacoustic emission input/output function modeling

Development and the role of internal noise in detection and discrimination thresholds with narrow band stimuli

Emily Buss,[a] Joseph W. Hall III, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

The experiments reported here examine the role of internal noise in the detection of a tone in narrow band noise and intensity discrimination for narrow band stimuli in school-aged children as compared to adults. Experiment 1 used 20-Hz wide bands of Gaussian and low-fluctuation noise centered at 500 Hz to assess the role of stimulus fluctuation in detection of a 500-Hz pure tone. Additional conditions tested whether performance was based on level and/or level-independent cues. Children's thresholds were elevated with respect to adults, and whereas adults benefited from the reduced fluctuation of low-fluctuation noise, children did not. Results from both groups were consistent with the use of a level cue. Experiment 2 estimated intensity increment thresholds for a narrow band Gaussian noise or a pure tone, either with or without a presentation-by-presentation level rove, an additional source of level variability. Stimulus variability was found to have a larger effect on performance of adults as compared to children, a rather counterintuitive finding if one thinks of children as more prone to informational masking introduced by stimulus variability. Both tone-in-noise and intensity discrimination data were consistent with the hypothesis that children's performance is limited by greater levels of internal noise. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354024]

I. INTRODUCTION

In many psychoacoustic paradigms, performance of young children is poor in comparison to the performance of adults. This has been found for relatively complex listening tasks, such as speech recognition under challenging listening conditions (e.g., Elliot, 1979; Nábelek and Robinson, 1982) and informational masking (Allen and Wightman, 1995; Hall *et al.*, 2005; Oh *et al.*, 2001; Wightman *et al.*, 2003). It has also been demonstrated for relatively simple tasks, such as the detection of a tone in quiet or in masking noise (Allen and Wightman, 1994; Fior, 1972; Maxon and Hochberg, 1982), frequency discrimination (Jensen and Neff, 1993; Maxon and Hochberg, 1982), and intensity discrimination (Fior, 1972; Maxon and Hochberg, 1982; see also Jensen and Neff, 1993). Although the time course of these developmental effects varies with paradigm and experimental conditions, there is at least some evidence of improvement out to 10–12 years of age in very simple psychoacoustic tasks (e.g., Maxon and Hochberg, 1982). A satisfactory explanation of these prolonged developmental effects for relatively simple psychoacoustic paradigms has proven elusive.

Previous research has considered a range of factors that might explain threshold elevation in children, including poor frequency selectivity, reduced motivation or inattention, and inefficient listening strategies. While any of these factors could contribute to the effects observed under some conditions, none seems to account for all effects observed. For example, Schneider *et al.* (1990) estimated critical bands and argued that the somewhat wider critical bands of infants and young children were not nearly wide enough to account for the elevation in masked thresholds observed. Motivation and inattention have been argued to be unlikely candidates based on relatively small effects of changing the task reward structure and the relative stability of thresholds over time for a given listener (for a review, see Schneider *et al.*, 1989). While listening strategy almost certainly plays a role in the performance of young children under some listening conditions, Willihnganz *et al.* (1997) provide evidence that it cannot account for all of the developmental effects observed, even under complex listening conditions. In that study, perceptual weights characterizing performance of adults and young school-aged children in a masked intensity discrimination paradigm were estimated; the pattern of weights did not suggest that children were using a particularly inefficient strategy when compared to adults, despite the finding of elevated thresholds for the younger listeners.

Schneider and colleagues (1989, 1992) suggest that increased internal noise might account for a wide range of developmental differences. While the phrase *internal noise* is sometimes used in a very broad sense, describing any of a number of inaccuracies that might account for deviations from optimal performance, the hypothesis proposed was more specific. In this case, internal noise was defined in terms of variability in the neural representation of intensity. Further, it was reasoned that increased internal noise would result in shallower psychometric functions in a masked detection task because energy at the output of the auditory filter centered on the signal would be more variable. Schneider *et al.* (1989) tested this hypothesis by estimating percent correct for the detection of a tone masked by a 1/3 octave band

of noise; psychometric functions from group data failed to show compelling evidence that children's psychometric functions are shallower than those of adults. However, other studies (e.g., Allen and Wightman, 1994; Olsho *et al.*, 1988) have estimated psychometric functions based on *individual* listeners' data and have found an increase in slope with age, as would be expected if internal noise played a role in the elevated masked detection thresholds of the child listeners.

Internal noise has been very useful in helping to account for various psychoacoustical data of adult listeners, including intensity discrimination (Jesteadt *et al.*, 2003). Although the hypothesis that developmental differences in threshold sensitivity can be accounted for in terms of differences in internal noise is attractive, it has not been widely applied to studies of the developmental effects observed in simple psychoacoustic tasks. Our main purpose in the present study was to test the hypothesis that developmental effects observed with a simple psychoacoustic task in school-aged children can be explained, at least in part, in terms of internal noise.

One general strategy that has been used previously to study internal noise in adult listeners is to observe the effects of manipulating the variability of some physical parameter of the stimulus, such as intensity, and observing the effects on listener performance (Jesteadt *et al.*, 2003; Spiegel and Green, 1981). In this approach, the magnitude of change in performance due to changes in the variability of the physical parameter is used to draw inferences about the neural variability (or internal noise) associated with the detection process. One type of variability commonly introduced in psychophysical paradigms is that of level rove, whereby a random intensity is selected prior to each presentation interval from a restricted range (cf. Jesteadt *et al.*, 2003). Another source of variability is the within-interval fluctuation of a stimulus (Spiegel and Green, 1981), such as that characterizing random noise samples. Both experiments reported here use the general approach of assessing performance in the face of stimulus variability and making inferences about the magnitude of internal noise based on susceptibility to external noise.

Narrow band Gaussian noise is characterized by relatively prominent amplitude fluctuation dictated by its bandwidth, as well as the relative magnitudes and phases of the constituent components. In this context, the addition of a pure tone signal produces a number of effects that could potentially be used as detection cues: the overall level of the stimulus is increased, but there are also changes in envelope statistics, temporal fine-structure regularity, and, possibly, spectral shape. Level-independent cues can be shown to support the detection of a tone in bandpass Gaussian noise under conditions where level is an unreliable cue (Richards, 1992; Kidd *et al.*, 1989; Richards and Nekrich, 1993), such as in the presence of an intensity rove, and may be combined with a level cue under more typical (nonroved) conditions (Richards, 2001). The cues available for a tone in "low-fluctuation noise," characterized by a relatively flat temporal envelope, are somewhat different from those available in Gaussian noise. Notably, adding a pure tone to a narrow band Gaussian

noise tends to flatten the amplitude envelope of the stimulus, while adding a tone to a narrow band low-fluctuation noise can increase envelope fluctuation.

In adult listeners detection thresholds for a pure tone added to a low-fluctuation masker tend to be lower than those for a pure tone added to a comparable bandpass Gaussian noise masker (Hall *et al.*, 1998; Hartmann and Pumplin, 1988; Kohlrausch *et al.*, 1997; Eddins and Barber, 1998; but also see Eddins, 2001). At least two possibilities have been suggested to account for this finding. Kohlrausch *et al.* (1997) suggested that the change in envelope statistics with the addition of a tonal signal is responsible. Eddins (2001) offered an alternative suggestion that the short-term stability of low-fluctuation noise maskers facilitate the detection of an increment in intensity. This interpretation was supported by data on intensity discrimination; thresholds for low-fluctuation noise were lower than those for Gaussian noise, even when the signal interval was associated with a change in level and no change in the envelope statistics. This interpretation is also consistent with the views of Bos and deBoer (1966), who argued that both intensity discrimination and masked tone detection are limited at narrow bandwidths (10–40 Hz) by inherent stimulus fluctuations.

As mentioned at the outset, intensity discrimination has been shown in some studies to be elevated in school-aged children. Thus, children might be poorer than adults in detecting a signal in noise because of relatively poor ability to utilize the associated level cue. It is less clear how well children make use of level-invariant cues for a tone added to narrow band noise. Allen *et al.* (1998) examined the types of cues used by children in a tone-in-noise detection task for a 400-Hz wide masker centered at 1 kHz. Preschool aged children were able to use level-invariant cues, such as temporal changes in the masker with the addition of a signal. It was also noted that at the detection threshold of children, multiple cues associated with the signal would be present in the physical stimulus: it was hypothesized that children may not be good at integrating the different cues available or switching between cues.

Very little is known about the effect of stimulus variability on young school-aged children in tone detection or intensity discrimination with narrow band stimuli other than tones. Preliminary data collected in our lab in the course of a binaural experiment, however, suggested that children 5–9 years of age might perform quite differently from adults in the face of stimulus variability. In that study (Buss *et al.*, 2003), diotic thresholds for a 500-Hz tone in a 20-Hz wide narrow band of noise centered on 500 Hz were measured for both Gaussian and low-fluctuation noise. While adults showed the expected masker effect, with lower thresholds in low-fluctuation noise, children had very similar thresholds in low-fluctuation and Gaussian noise, suggesting that they may not be able to take advantage of the cues available in low-fluctuation noise. Experiment 1 pursued this preliminary finding by estimating monaural pure-tone detection thresholds in both Gaussian noise and low-fluctuation noise. Additional conditions measured sensitivity to either level cues or level-independent cues present in the tone-in-noise task in order to confirm that any age effects obtained with a tone in

noise were related to level (and not level-independent) cues. Experiment 2 used stimuli with high or low degrees of inherent fluctuation and examined the effects of additional variability introduced by level rove on intensity discrimination, a paradigm closely resembling that used in adults by Jesteadt *et al.* (2003) to estimate internal noise.

II. EXPERIMENT 1: ENVELOPE FLUCTUATION AND MONAURAL CUES FOR MASKED TONE DETECTION

The purpose of Experiment 1 was to test the prediction that if internal noise is the limiting factor in children's detection of a tone in narrow band noise, then they should not benefit from reduced masker fluctuation to the same extent as adults, for whom external noise plays more of a role in performance. In other words, elevated levels of internal noise could result in a failure of children to make effective use of the more stable level cues in low-fluctuation as compared to Gaussian noise. Because this logic rests on the assumption that listeners are using a level cue to perform the task, and because multiple cues are available for these stimulus conditions, additional conditions were run in which level and level-independent cues were assessed separately. Results of these additional conditions will be interpreted in terms of the cues most likely to underlie performance in the tone-in-noise conditions. Further, if children have difficulties integrating multiple cues in the basic conditions, as suggested by the results of Allen *et al.* (1998), then these blocks of stimuli characterized by reduced cues (e.g., just level) should give a more accurate indication of the child listeners' abilities to make use of level and level-independent cues.

A. Methods

1. Listeners

A group of 8 children participated, including 6 males and 2 females, aged 5.0 to 10.5 yr (mean=7.3 yr). The adult group was comprised of 10 listeners, 2 males and 8 females, aged 18.0 to 48.0 yr (mean=28.4 yr). All listeners had pure tone thresholds equal to or better than 15 dB HL for octave frequencies 250 to 8000 Hz (ANSI, 1996). None of the listeners reported a history of ear problems. A history of ear disease, including chronic otitis media (OM), or an active case of OM at the time of testing were considered grounds for exclusion from the study. Data for one additional 7-years-old child listener were excluded from the study due to excessive variability in performance. Thresholds for this child varied by more than 25 dB in one condition in a single session: this variability was attributed to fatigue and inattention rather than to learning because thresholds increased (worsened) over the course of the session.

2. Stimuli

The Gaussian masker was a 20-Hz wide band of noise centered on 500 Hz. The low-fluctuation noise was generated following a procedure described by Kohlrausch *et al.* (1997). First, a 20-Hz wide band of Gaussian noise centered on 500 Hz was generated; the sample was divided by its Hilbert envelope in the time domain and then restricted to the origi-

nal 20-Hz spectral region via multiplication in the frequency domain, a procedure that was repeated 10 times for each sample. This process resulted in a masker with a relatively flat temporal envelope. Quantifying envelope fluctuation as the ratio between the standard deviation and the mean of the envelope (V), the low-fluctuation and Gaussian noise maskers are characterized by $V=-24.1$ and $V=-5.6$ dB, respectively. In terms of the modulation depth (m) for sinusoidal amplitude modulation, these values are comparable to $m=0.09$ and $m=0.74$, respectively (see Kohlrausch *et al.*, 1997). Maskers were scaled to a digital amplitude associated with 65 dB SPL and presented for 409-ms, including 50-ms $\cos^2$ ramps. The signal was a 500-Hz pure tone, 309 ms in duration. When present, the signal was ramped on and off using 50-ms $\cos^2$ ramps, with onset occurring 100-ms after onset of the masker, and the signal level was adjusted by way of a scalar. Stimuli were generated digitally in MATLAB and played out via two channels of a DAC (TDT, RP2) at approximately 12.2 kHz. The resultant streams were routed through a headphone buffer (TDT, HB7) to the left earphone of a pair of circumaural earphones (Sennheiser, HD 265).

Maskers were generated using arrays of 2^{13} points. The effect of truncating these arrays to the number of points associated with 409 ms (4993 points) and imposing ramps introduced some variability around the 65-dB mean masker level. This effect was estimated based on 100 samples, generated following the procedures described above, with the level for each sample calculated based on the steady portion of the stimulus (i.e., excluding ramps). For Gaussian noise stimuli, 50% of the samples were within ±0.5 dB of 65 dB SPL, 75% were within ±1 dB, and 90% were within ±1.5 dB. For low-fluctuation noise stimuli, 90% of the samples were within ±0.5 dB of 65 dB SPL. The effects of this source of stimulus variability will be revisited below.

For the purposes of stimulus generation, the 409-ms masker was divided into two portions: the *fringe* and the *core*. The *fringe* portion of the masker was 150-ms in duration, including 50-ms $\cos^2$ onset and offset ramps. The *core* portion of the masker received the same gating as the pure tone signal: that is, it was 309-ms in duration, including 50-ms $\cos^2$ ramps. The *fringe* offset and the *core* onset masker ramps overlapped in time, such that the masker sounded continuous when the two portions were played together. Dividing the masker in this way allowed the independent manipulation of the *core* portion of the masker in conditions where cues were limited to either just level or just envelope. The primary motivation for the use of a masker fringe was to allow comparison with tone-in-noise data collected using an asynchronous onset. The implications for adopting the *core/fringe* stimulus configuration in the present study are considered further in the discussion section.

In the *all-cues* condition, the 500-Hz pure-tone signal was simply added to the masker with no further adjustment made to the *core* portion of the masker. This condition is referred to as the *all-cues* condition because the addition of the signal introduces both level and level-independent cues. In a second condition, the signal was an increment in intensity of the *core* portion of the masker. In this case, waveform amplitude of the *core* portion of the masker was adjusted

according to the increment that *would have been obtained* by the addition of a pure tone signal at the associated signal level. This was computed with the formula $10 \log(10^{65/10} + 10^{\text{Sig Lev}/10})$, where Sig/Lev is the level of the simulated pure-tone signal, in dB. As such, thresholds in this condition can be directly compared to those obtained in the *all-cues* condition. This condition is referred to as the *level-only* condition because the signal consists of an increment in the level in the core portion of the masker and is not associated with level-independent cues, such as a change in envelope statistics within that portion of the masker. In the final condition, a pure tone signal was added to the masker, but the sum (signal+core) was scaled down to counteract any resultant change in level. In this *no-level* condition there was a change in the stimulus envelope with the addition of the signal, but no change in the overall level. Other possible cues in the *no-level* condition include the regularity of zero-crossings and spectral shape, although the latter was unlikely to be a viable detection cue for such a narrow bandwidth.

Maximum values for *no-level* tracks were imposed because of the nonmonotonicity of some of the temporal stimulus features thought to provide possible cues. For example, the temporal envelope of a low-fluctuation noise masker alone is relatively flat. With the addition of a tone, envelope variability increases with increases in signal level up to a point, but beyond that point variability begins to drop, as the stimulus envelope becomes dominated by the pure-tone signal. Based on these considerations, and calculations of possible level-invariant cues (i.e., average envelope slope, envelope max/min ratio, and periodicity of zero crossings), it was decided to impose a track ceiling value of 65 dB SPL for low-fluctuation noise, the approximate signal level associated with the most prominent level-invariant cues. A ceiling of 90 dB SPL was set for the *no-level* Gaussian noise masker stimuli; while level-invariant cues were monotonic for this masker type, it was reasoned that failure to make use of these cues with a 90 dB signal would not likely change in the face of further increases in signal strength. A threshold estimate was judged to be at ceiling if the threshold estimation track hit ceiling three or more times during a run, and an average data point for a given condition was judged to be unmeasurable if two or more threshold estimates were at ceiling.

To summarize, there were six conditions in total, illustrated in Fig. 1. In this figure, an example Gaussian noise sample is shown in the left panel and an example low-fluctuation noise sample in the right. The repeated gray waveform is the masker alone and the associated black lines indicate the envelope of a signal-plus-masker, with both signal and masker presented at a level of 65 dB SPL. At the top of each panel are indications of the temporal placement (and overlap) of the *fringe* and *core* portions of the stimuli. In the *all-cues* condition the signal was a pure tone gated with the ramps defining the *core* portion of the stimulus; interactions between signal and masker produced both level and envelope cues. In the *level-only* conditions the signal consists of an increment in masker level, defined as the increment that would have been produced if a pure tone signal had been added. In these conditions there was no change in envelope statistics over the course of the signal presentation to cue the
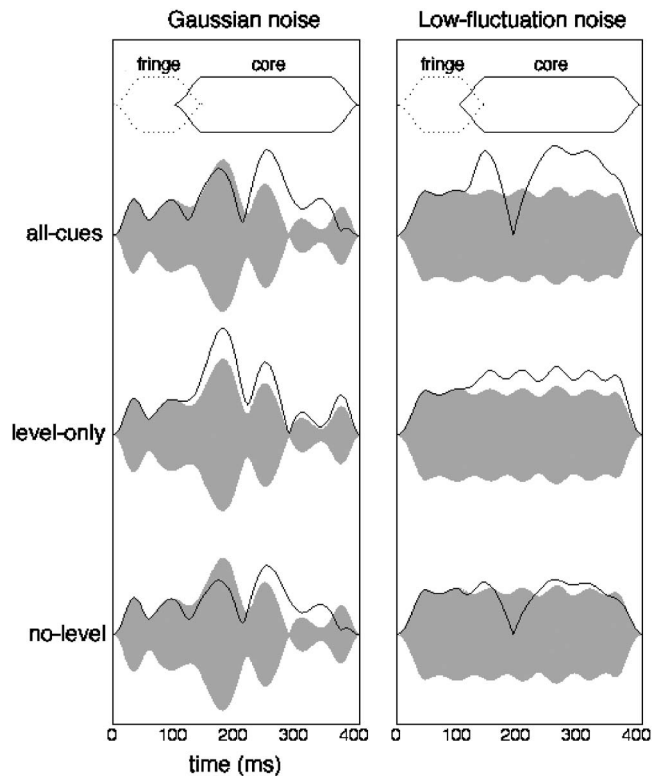
FIG. 1. Example stimuli from Experiment 1 are shown, with Gaussian noise (left panel) and low-fluctuation noise (right panel), and signal conditions indicated in the left margin. The top row in each panel shows the segments of the stimulus identified as the *fringe* and core portions. The next row, labeled *all cues*, shows a standard stimulus (in gray) and the envelope of that standard summed with a 65-dB SPL pure tone signal (black line). The final two rows show standard and signal-plus-standard for the comparable *level-only* and *no-level* conditions. For illustrative purposes a single random sample was selected for all three signal conditions within each panel; in the experiment a new random sample was generated prior to each presentation.

presence of the signal other than those at the boundary of the *fringe* and *core* segments. In the *no-level* conditions the overall level of the *core* portion of the stimulus was scaled back after addition of the signal such that a change in envelope statistics over the course of the signal presentation cued the presence of the signal, but there was no increase in level.

3. Procedures

Threshold estimates were obtained using a 3-alternative forced-choice procedure and a 3-down 1-up tracking rule, estimating 79% correct on the psychometric function (Levitt, 1971). Initial signal level adjustments were made in steps of 4 dB, and level adjustments were reduced to 2 dB after the second track reversal. Each track continued for a total of six reversals, and the final threshold estimate was the mean level at the last four reversals. Listening intervals were marked visually via animation on a computer screen. Over the course of a track a cartoon picture was unmasked, in the style of a jigsaw puzzle, with one piece revealed following each correct response. No visual feedback was provided, following an incorrect response. At the end of the track the cartoon was fully revealed and performed a two-second animation. All listeners used this interface.

On the first visit to the lab, child listeners were randomly assigned to one of two groups. One group ran a block of the *all-cues* and a block of the *level-only* condition for each masker type. The other group ran a block of the *all-cues* and a block of the *no-level* condition for each masker type. On the second visit to the lab, child listeners performed a second block of the *all-cues* condition and a block of the reduced-cue condition they had not previously listened to (either *level-only* or *no-level*) for each masker type. Within a visit, the order of the four blocks (2 signal conditions × 2 masker types) was randomized. Adult listeners followed the same sequence of blocks, but were able to complete more than four blocks in a visit. Each visit lasted no longer than one hour, with frequent breaks. A block of trials consisted of three threshold estimates, with a fourth if the span in initial estimates was 3 dB or more. All estimates for a listener were averaged to produce the final thresholds reported below.

Prior to each block, the experimenter described the detection cue associated with that condition. For the *level-only* condition, listeners were asked to select the loudest interval. In the *no-level* condition, the signal was described as the "more wobbly" sound for low-fluctuation noise and the "smoother, or less wobbly" sound for Gaussian noise. In the *all-cues* condition, instructions were to listen to loudness and/or degree of fluctuation. In all cases, use of visual feedback to optimize performance was encouraged.

B. Results

The *all-cues* condition was performed twice by each listener for each masker type. This was done in part to assess the stability of this very important baseline condition. To that end the first and the second replicate thresholds were compared for evidence of improvement.[1] One-tailed, paired *t* tests revealed no improvement between the first and second replication for Gaussian ($t_{16}=-1.16$, $p=0.13$) or low-fluctuation noise ($t_{16}=0.90$, $p=0.19$). Tests with just child and just adult data similarly resulted in no significant difference between the first and second block of thresholds ($\alpha=0.05$). Subsequent analyses incorporating *all-cues* data were performed based on the first set of estimates obtained, although results are qualitatively identical when analyses are performed on the mean of all (six to eight) threshold estimates in the *all-cues* condition.

Figure 2 shows the mean thresholds for each group (bars) and individual listeners' thresholds (circles) obtained with Gaussian noise (top) and low-fluctuation noise (bottom). Individuals' thresholds are ordered by listener age, youngest (left) to oldest (right) within group. Open bars and open circles indicate adult data, and hatched bars and filled circles indicate child data. In the *all-cues* condition, mean thresholds for adults were 68.2 and 62.5 dB SPL for Gaussian and low-fluctuation noise, respectively. This difference of 5.7 dB was significantly greater than zero in a paired *t* test ($t_9=5.09$, $p<0.001$, two-tailed). This result is comparable to the 5-dB effect reported by Hall *et al.* (1998) for 10-Hz wide noise bands centered at 500 Hz. In contrast, for children the mean thresholds were 69.9 and 70.4 dB in Gaussian and low-fluctuation noise. The masker effect of
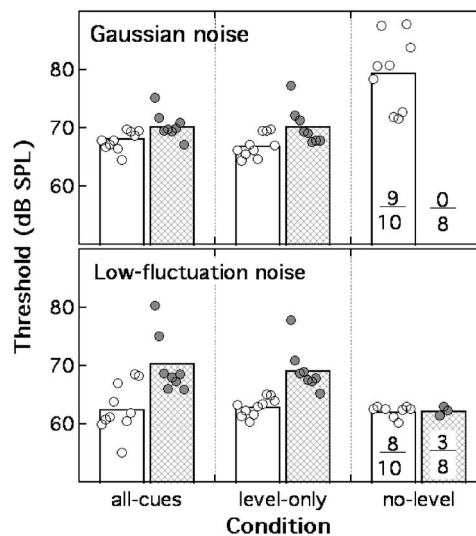
FIG. 2. The mean thresholds from Experiment 1 are plotted for Gaussian noise (top panel) and low-fluctuation noise (bottom panel). Bar markings indicate mean data from adults (□) and children (⊠), and the circles associated with each bar show individual data. Stimulus conditions are indicated along the abscissa. The ratios at the right of each panel indicate the proportion of listeners contributing to each mean in the *no-level* conditions: individuals' data were omitted from the average if they were deemed to be indistinguishable from ceiling performance.

only 0.5 dB was not significant ($t_7=0.105$, $p=0.92$, two-tailed). This result closely replicates results for analogous diotic stimuli collected in the course of a binaural study (Buss *et al.*, 2003).[2] The +5 dB S/N ratio at threshold for the Gaussian noise masker can be compared with the results of Hall *et al.* (1997). In one condition of that study, thresholds were estimated for the detection of a 500-Hz pure tone masked by a 20-Hz wide band of noise centered on 500 Hz. The masker was 600 ms in duration, and the signal was 400 ms, with signal onset occurring 200 ms after masker onset. The average threshold for school-aged listeners in that study (ranging from 5–11 yrs) was +3.4 dB S/N, comparable to the +5 dB S/N observed here.

The second set of bars in each panel of Fig. 2 shows thresholds in the *level-only* condition. In all cases these thresholds are quite similar to those in the corresponding *all-cues* condition. A repeated-measures ANOVA was performed to assess this relationship. There was one between-subjects factor (AGE: child, adult) and two within-subject factors (CUE: *all-cues*, *level-only*; MASKER: Gaussian, low-fluctuation). There was a significant main effect of AGE ($F_{1,16}=16.23$, $p<0.001$) and a significant main effect of MASKER ($F_{1,16}=40.42$, $p<0.0005$). The interaction between AGE and MASKER was significant ($F_{1,16}=27.77$, $p<0.0005$). There was no main effect of CUE ($F_{1,16}=1.56$, $p=0.23$) and no interactions between CUE and the other factors. This confirms that the *level-only* and *all-cues* thresholds were not statistically different. Thus, there was a significant masker effect in the *level-only* condition for adults, but no significant effect for children.

The right-most set of bars in each panel of Fig. 2 shows results of the *no-level* condition. Only those data points deemed to be measurable were included in the average, so these results tend to underestimate thresholds by way of ex-

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Buss *et al.*: Development and internal noise   2781

cluding large values at ceiling. The proportion of data points included in each mean is indicated at the base of each bar or in place of a bar. For the Gaussian noise masker, none of the children obtained a threshold below the ceiling value of 90 dB SPL. All but one of the adults performed below the ceiling, but thresholds in this condition were elevated more than 10 dB relative to the *all-cues* condition. These results suggest that cues other than level (e.g., envelope) probably played little or no role in performance in the *all-cues* condition with Gaussian noise. For low-fluctuation noise, 3/8 of children and 8/10 of adults produced a threshold estimate below the ceiling of 65 dB SPL. In all cases where a measurable threshold was obtained, the resultant threshold was near 62 dB, similar to average adult *all-cue* threshold and 8 dB better than average child *all-cue* threshold. The three children who performed below ceiling in the *no-level* condition attained thresholds approximately 5 dB lower than their thresholds in the associated *all-cue* condition. This suggests that these listeners failed to make use of level-independent cues present in the *all-cue* condition.

C. Discussion

Experiment 1 showed that adults were more sensitive at detecting a tone added to a low-fluctuation masker as compared to a Gaussian noise masker, and children showed no such masker effect. While children's detection thresholds did not vary as a function of masker type, it seems reasonable to assume that children were able to perceive a difference between Gaussian and low-fluctuation noise stimuli. A 20-Hz wide band of Gaussian noise has an equivalent AM rate of approximately 13 Hz (Rice, 1953). Based on the temporal modulation transfer function (TMTF) data of Hall and Grose (1994), at a modulation rate of near 13 Hz, AM of the Gaussian noise (with $m=0.74$) should be clearly perceptible for all child listeners, while modulation for the low-fluctuation stimuli (with $m=0.09$) would be expected to fall below the modulation detection threshold. This suggests that child listeners in Experiment 1 should be able to perceive the difference between the envelope fluctuation of Gaussian and low-fluctuation noise, and yet these perceptible differences did not affect the performance. There are no published AM discrimination data that permit the assessment of the relative sensitivity to AM in children and adults, another factor that could bear on the absence of a masker effect in the current dataset.

It is possible that children might have particularly elevated thresholds for a tone masked by a low-fluctuation noise when compared to adults because of increased cue complexity. If the cues allowing adults to achieve lower thresholds with low-fluctuation noise are more subtle or complex than those present in the Gaussian noise case, and if children are not as adept as adults at combining or switching between cues available in low-fluctuation noise, then low-fluctuation noise thresholds in children would not benefit from these additional subtle or complex cues. This hypothesis is similar to that proposed by Allen *et al.* (1998) to account for tone-in-noise masking results obtained with preschool-aged children. The fact that 3 out of 8 child listen-

ers attained lower thresholds in the *no-level* than the *all-cues* conditions is consistent with this view.

Results of the reduced-cue conditions suggest that an increment in level may be the primary cue underlying thresholds for a tone added to a narrow band of noise for most listeners, both children and adults. Level-invariant cues in isolation were associated with elevated or unmeasureable thresholds for the Gaussian noise masker, and so almost certainly played no material role in detection for the *all-cues* Gaussian condition. The level-invariant cues associated with the low-fluctuation masker support a relatively good performance for some adult listeners, comparable to that seen in the *all-cues* condition. The failure of many listeners to attain a threshold below the ceiling value of 65 dB SPL, however, suggests that this cue is not viable for all adult listeners and fewer than half of the child listeners. These observations are consistent with the conclusion that performance in the *all-cues* condition can be most parsimoniously explained in terms of the change in level across intervals associate with the addition of a signal. This does not imply that child observers are not able to make use of level-invariant cues, in general. Indeed, Allen *et al.* (1998) showed that preschool listeners are able to make use of level-invariant cues in the detection of a tone in bandpass noise spanning 800–1200 Hz. The assertion is just that the *present* results are likely based on level.

One aspect of the current paradigm that varies from the one employed by Allen *et al.* (1998) is the use of a leading *fringe*. It has been shown that asynchronous onset can aid in sound source segregation (Bregman, 1990), and that school-aged children can make use of gating cues in an informational masking paradigm (Hall *et al.*, 2005). Based on these findings it is plausible that delaying the onset of the signal portion of the stimulus could have the benefit of improving signal detection by virtue of improved segregation. A second factor of interest is the role of memory in this paradigm. The presence of a fringe prior to signal presentation offers the opportunity for the subject to make a within-interval comparison of the *fringe* and *core* segments of the stimulus. Because no across-interval comparison is necessary for this strategy, the availability of such a cue could improve the performance of child listeners. Alternatively, asynchronous onset could introduce added complexity and make the task more difficult for child listeners just by virtue of increasing the number and quick succession of different stimulus features. This is unlikely in light of the results of Hall *et al.* (1997), however. That study measured thresholds for school-aged and adult listeners for stimuli comparable to those in the *all-cues* condition. No difference was observed between a condition in which a narrow band masker was played continuously and a gated condition, where the masker was gated on 200 ms prior to the signal and both the signal and masker were gated off synchronously. This result argues against the idea that increased task complexity associated with the fringe could have elevated thresholds in the present study, at least for Gaussian noise conditions. One motivation for Experiment 2 was to test the fidelity with which level cues can be used under conditions of synchronous onset to determine

whether the results of Experiment 1 would generalize to other stimulus conditions.

One interesting result of Experiment 1 is the significant MASKER $\times$ AGE interaction in the *level-only* condition. As in the *all-cues* condition, thresholds in the *level-only* condition varied across masker for adults but not for children: adults' thresholds in the *level-only* condition were 4.1 dB lower for low-fluctuation than Gaussian noise, while for children the difference was a (nonsignificant) 1.0 dB. The advantage for intensity discrimination in low-fluctuation as compared to Gaussian noise obtained for adults was noted by Eddins (2001). In that study adults discriminated level across interval for 50-Hz wide bands of noise, and average thresholds were 3.2 dB lower for low noise than Gaussian noise. Eddins speculated that the reason for the different intensity discrimination thresholds across stimulus types was the stability of the short-term level cue. However, if the performance of children is limited by internal noise rather than external noise, as hypothesized in the Introduction, then children would not be able to benefit from this reduction in external noise. Results of Experiment 1 provide preliminary support for this hypothesis.

III. EXPERIMENT 2: THE EFFECT OF STIMULUS VARIABILITY ON INTENSITY DISCRIMINATION

Whereas the paradigm of Experiment 1 examined the effects of inherent stimulus fluctuation on the use of level cues in adults and children, Experiment 2 incorporates level rove as an additional source of stimulus variability. Roving the level of the stimuli on an interval-by-interval basis introduces external noise that would be expected to elevate thresholds to the extent that external noise (as opposed to internal noise) imposes limits on performance. Similar to the approach taken in Experiment 1, the hypothesis of Experiment 2 is that children's intensity discrimination thresholds should be less severely elevated by level rove than those of adults. This result would be expected if the performance of children is limited by internal noise to a greater extent than that of adults.

To that end, Experiment 2 addresses the role of both within-interval and across-interval stimulus variability in intensity discrimination of adults and children. Thresholds for detecting an across-interval increment in intensity were estimated for a narrow band Gaussian masker and for a pure tone, characterized by pronounced variability and very steady within-interval intensity, respectively. An additional manipulation imposed a random level rove to each stimulus presentation, which has the consequence of increasing the variability of the intensity cue.

A. Methods

1. Listeners

A group of 15 children participated, including 10 males and 5 females, ages 5.0 to 10.5 yr (mean=7.9 yr). The adult group was comprised of 12 listeners, 3 males and 9 females, ages 17.1 to 50.0 yr (mean=29.2 yr). Exclusion criteria and audiometric status were the same as those described for Experiment 1. Two additional child listeners (5 and 7 yr) were excused due to excessive variability in threshold estimates; these listeners consistently provided a 10+ dB spread in thresholds within condition. Two of the child listeners and two of the adult listeners had previously participated in Experiment 1.

2. Stimuli

Stimulus generation was similar to that described for Experiment 1. Stimuli were either 20-Hz wide bands of Gaussian noise centered on 500-Hz or 500-Hz pure tones, ramped on and off with 50-ms $\cos^2$ ramps and 409-ms in duration (including ramps). Noise samples were generated in the time domain, converted to the frequency domain for filtering by way of multiplication with a boxcar function, converted to the time domain, truncated to 409 ms and ramped with 50-ms $\cos^2$ functions. Pure tones were generated in the time domain, truncated and ramped. Like the *level-only* condition of Experiment 1, the standard interval stimulus was 65 dB SPL and the signal was an increment in the level, produced by adding a scaled copy of the standard. In contrast to Experiment 1, such increments spanned the entire (409-ms) stimulus duration. In an additional manipulation, the standard presented in each interval was roved in level, according to a random draw from a uniform distribution ±6 dB. The level of the "scaled copy" (producing the level increment in the signal interval) was not roved in these conditions.

The condition in which the tonal standard was roved in level will be referred to as a *tone roved*, and the condition with no rove will be referred to as the *tone-steady* condition. Analogous conditions with Gaussian noise stimuli will be referred to as *Gnoise-roved* and *Gnoise-steady*, respectively.

3. Procedures

As in Experiment 1, listeners performed a 3-alternative forced-choice track with visual indications of the listening intervals and positive feedback following a correct response. Instructions were to select the loudest of the three intervals. The signal level was adjusted in a 3-down, 1-up procedure, with 4 dB steps at the outset and 2 dB after the first 2 reversals. As above, thresholds are reported *as if* the signal was added to the masker in random phase, to facilitate comparison of these results with the tone-in-noise results of Experiment 1. The four conditions were run in blocks, with blocks visited in random order. Three threshold estimates were obtained in each block, with a fourth collected in cases when the initial three spanned a range of 3 dB or more. Listening sessions lasted no more than 1 hour, with frequent breaks. In most cases child listeners were able to complete all four conditions in this time period, although in several cases more than one test session was required. All adult listeners completed the study in a single 1 h session.

Thresholds in this experiment were highly variable, particularly in the child group. Effects of this variability were counteracted by the following procedure for identifying outliers. The difference between the maximum and the minimum threshold in a block of estimates was computed. If that value was 6 dB or greater, the set was visually examined for
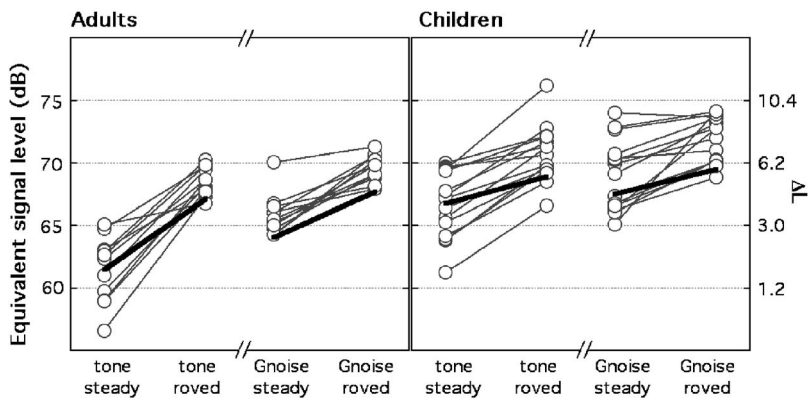
FIG. 3. Individual data from Experiment 2 are shown in units of dB *as if* the signal had been added in random phase. The ordinate labels at the right show associated units of ΔL. Panels show results for adults (left panel) and children (right panel), with stimulus conditions indicated along the abscissa. Dark lines indicate fits to the data.

outliers, defined as a single threshold estimate that is at least 4 dB greater than or less than all the other estimates in that condition. If an outlier was identified, it was removed from the mean. Following this procedure, fewer than 5% of the child data and 2.5% of the adult data were culled from the results presented below.[3]

B. Results

Thresholds for each listener in each condition are indicated with open circles in Fig. 3, with data for adults in the left panel and those for children in the right panel. The labels on the abscissa indicate the associated stimulus condition. The left ordinate indicates signal level at threshold in dB and the right ordinate shows associated units of ΔL, to facilitate a comparison with published intensity discrimination data. Average thresholds in the *Gnoise-steady* condition are 65.8 dB for adults and 68.9 for children. These values are within 2.5 dB of those observed in comparable conditions in Experiment 1. The inclusion of stimulus rove increased these thresholds by 3.7 for adults and by 3.1 dB for children. Average thresholds in the *tone-steady* condition are 61.5 for adults and 66.7 for children. The inclusion of rove increased these thresholds by 7.0 and 4.0 dB for the adults and children, respectively. Thus, the effect of rove was larger for the pure tone than for the noise stimulus, and the effect of rove for the pure-tone stimulus was larger for adults than for children. Comparing *tone-steady* and *Gnoise-steady* conditions, the effect of stimulus type was 4.3 dB for adults and 2.2 dB for children. Thus, increases in threshold associated with introduction of stimulus variability, either rove or inherent (within interval) fluctuation, were larger for adults than children when using the *tone-steady* condition as a baseline.

A repeated-measures ANOVA was performed on data in units of dB[4] to assess the significance of these effects. There was one between-subjects factor (AGE: adult, child) and two within-subject factors (ROVE: steady, roved; STIM: Gaussian noise, tone). There were significant main effects of AGE $(F_{1,25}=24.97, p<0.0001)$, ROVE $(F_{1,25}=187.98, p<0.0001)$ and STIM $(F_{1,25}=53.88, p<0.0001)$. The two-way interaction between ROVE and STIM was significant $(F_{1,25}=19.93, p<0.001)$, reflecting a larger effect of rove for the tone as compared to Gaussian noise. The ROVE$\times$AGE effect $(F_{1,25}=7.89, p<0.01)$ reflects greater susceptibility to the rove of adults as compared to children. The STIM $\times$AGE interaction failed to reach significance $(F_{1,25}$

$=2.14, p=0.16)$. The significant three-way interaction $(F_{1,25}=6.12, p<0.05)$ is consistent with the observation that the largest rove effect is associated with the *tone-steady* data of adults.

Comparing the results of the *all-cues* condition of Experiment 1 to those of the no-rove conditions of Experiment 2 indicates analogous effects of stimulus fluctuation. In Experiment 1, thresholds of adult listeners dropped (improved) by 5.7 dB, comparing Gaussian with low-fluctuation noise conditions. In Experiment 2 a drop of 4.3 dB is observed comparing the *Gnoise-steady* and *tone-steady* conditions. For child listeners, improvements in threshold were smaller than those observed in adults, with values of 0.5 and 2.3 dB, respectively. These results suggest that the *fringe/core* stimulus configuration of Experiment 1 was not of great consequence to the interaction between age and effects of stimulus fluctuation.

C. Discussion

The results of Experiment 2 suggest that stimulus variability had less of an effect on the results of child than adult listeners, consistent with the hypothesis that child listeners are operating under conditions of elevated internal noise. One relatively uninteresting source of error in child listeners' performance is confusion regarding the signal cue. As noted by Jesteadt *et al.* (2003), feedback can often be misleading in roved intensity discrimination. Under such conditions the stimulus in the "signal" interval can, in fact, be the least intense; this occurs when the "signal" interval is associated with a low value of rove and/or the "no-signal" intervals are associated with a high value of rove. It is plausible that children could be more prone to confusion in the face of spurious feedback than adults. There are several considerations suggesting that this probably did not play a substantial role in the results obtained here, however. First, both adults and children performed the experiment using an interface where correct responses were rewarded by the unmasking of one piece of a "puzzle." Following an incorrect response there was no such unmasking. Listeners were not informed explicitly about this reward structure, and it seems likely that positive feedback (unmasking of a piece) would be more salient than negative feedback (nothing happening). As such, the listener would not receive salient feedback that an unselected and less intense stimulus was in fact the signal. Second, if spurious feedback causes more confusion in children than

adults, then one might expect to see a larger effect of rove for children than adults. In fact, the opposite result was observed, with child listeners' data more closely resembling that of adults in the roved conditions. Finally, there is a good correspondence in the pattern of results obtained in Experiments 1 and 2. In Experiment 1, external variability was manipulated by way of stimuli with large and small fluctuations over time, but no interval-by-interval rove (and hence no spurious feedback). As such, the most parsimonious explanation of the results of Experiment 2 would not rely on confusion associated with rove. If the elevated thresholds of the child listeners are due to some factor that is not specific to the current paradigm, then, the challenge is how to characterize that error and identify its source.

One way to summarize the findings of Experiment 2 is to note that the largest effect of stimulus variability is observed in cases where the baseline sensitivity was quite good (e.g., adult, *tone-steady* data). This observation prompted an attempt to model these results in terms of the combined effects of internal and external noise. In classic treatments of internal noise, sensitivity (d') is defined in terms of the mean difference in signal and no-signal cues (Δ) divided by the standard deviation of those underlying cue distributions (σ). The value of σ can be decomposed into internal noise (σ_i) and external noise (σ_e), resulting in the equation

$$d' = \Delta / \sqrt{(\sigma_e^2 + \sigma_i^2)}. \tag{1}$$

Jesteadt *et al.* (2003) provide a modern treatment of this classical approach. While there is some debate regarding the appropriate units in which to perform these calculations, following Jesteadt *et al.*, calculations reported here were performed in units of ΔL, and internal and external variability was assumed to add in dB.

The 3-down 1-up tracking procedure used in Experiment 2 estimates 79% correct, associated with a d' of 1.61 for a 3-alternative forced-choice task. Assuming that thresholds in the *tone-steady* condition are dominated by internal noise (i.e., $\sigma_e \approx 0$), the value of internal noise can be estimated as $\sigma_i = \Delta / 1.61$. For adults' mean data, this produces an estimate of internal noise of $\sigma_i = 0.99$ comparable to that reported for the adult listeners of Jesteadt *et al.* (2003). For children, the estimate of internal noise was considerably larger, with $\sigma_i = 2.46$.

For the remaining stimulus conditions, external noise is greater than zero, due to rove and/or small differences in level that arose because stimuli were equalized based on a longer sample than the one presented. Procedures for estimating thresholds based on these sources of variability are discussed in Appendix A. Predicted values of ΔL appear as dark lines in Fig. 3. These estimates capture the general trends in the data, most notably the biggest effects of external noise seen when comparing adult *tone-steady* to *tone-roved* conditions. Predicted thresholds in the *tone-steady* condition are quite close to the mean observed thresholds in each group; this is a consequence of the fact that values of σ_i were fitted based on these thresholds. Predicted thresholds in other conditions tend to underpredict performance by 1.4–2.4 dB. In adult data, predicted thresholds fall 3+ standard errors of the mean below the mean in all three condi-

tions characterized by high external noise. A similar pattern is observed in the data of child listeners, with the caveat that increased variance across individuals, reduces the magnitude of differences relative to estimates of the standard error. The underprediction of thresholds in cases of elevated values of σ_e is consistent with the findings of Eddins (2001), who showed that intensity discrimination is impaired for fluctuating (Gaussian noise) as compared to relatively stable (low-fluctuation noise) stimuli, even though overall stimulus level was stable in both cases. The fact that both rove and inherent fluctuation are associated with higher-than-expected thresholds in the current dataset suggests that these results are probably not due to the effects of amplitude modulation on loudness (e.g., Zhang and Zeng, 1997). Regardless of the source, this effect is roughly comparable in adult and child listeners' data, with an average mismatch of 1.7 and 1.9 dB between the prediction and mean threshold, respectively.

These discrepancies aside, the results shown in Fig. 3 suggest that the differential effects of stimulus variability across subject groups are broadly consistent with a very simple model of the combination of internal and external noise. By this account, children are less susceptible to the effects of stimulus variability because their rather large internal noise has a dominant effect on the internal representation of the cue. Similar logic can be used to account (at least in part) for the pattern of results obtained in the *all-cues and level-only* conditions of Experiment 1. In these data, adults' thresholds were more affected by the degree of inherent masker fluctuation (Gaussian versus low-fluctuation noise) than were the children's thresholds. This result is consistent with the argument that the lower levels of internal noise of adults make them more susceptible to the disruptive effects of external noise (i.e., stimulus variability).

While this approach is reasonably successful in characterizing the general trends in processing errors of child as compared to adult observers, it does not identify the source of that error. The foundation of the approach taken here (and in Jesteadt *et al.*, 2003) rests on the assumption that the source of noise reflects the physiological limitations associated with the encoding of the cues underlying performance. Self-generated noise, such as that produced by breathing or shifting position, could introduce noise that is not specific to the auditory system and might well be better characterized as "external" from the point of view of auditory processing. Further, the model proposed here assumes that the cues underlying performance are submitted to an optimal detector to generate a response, a process that is well characterized by signal detection theory. Evidence that children behave according to the rules of signal detection theory is scant, however. In the classic model, for example, the psychometric function has a well-defined shape determined by the internal noise distribution. It is possible that child listeners may change strategy or lose motivation under conditions of increased task difficulty, rather than consistently following a well-specified strategy, a result that could be reflected in an oddly shaped psychometric function. If the expectations of the classic model are met, however, the larger estimate of internal noise in child listeners predicts a shallower psychometric function for intensity discrimination. Work currently

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Buss *et al.*: Development and internal noise   2785
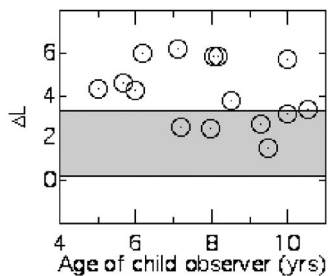
FIG. 4. Individual data of child listeners in the *tone-steady* condition are plotted as a function of age. Consistent with units used to compute internal noise, data are plotted in units of ΔL, with ±2 std around the mean adult threshold indicated as the shaded region.

underway will address the degree to which internal noise in child listeners, as estimated here, is consistent with the principals of signal detection theory.

Whereas it is not uncommon to consider internal noise as a qualitative factor in children's psychophysical data, it is not commonly quantified, as it was for these data. One exception is a study by Allen and Nelles (1996). That study estimated internal noise in adults and children for a frequency discrimination task and reported that internal noise was elevated for young children and reached levels comparable to that observed in adults by age 7. Because internal noise for the results of Experiment 2 was estimated based on ΔL in the *tone-steady* condition (that is, $\Delta L/1.61$), the effects of age on estimates of internal noise can be assessed directly based on those data. Figure 4 shows ΔL for individual child listeners in the *tone-steady* condition plotted as a function of age, with ±2 standard deviations around the mean adult threshold indicated in the shaded region. Here 10 out of the 15 child listeners fall above the 95% confidence interval for the adult performance. Those who fell within the confidence interval were, on average, older than the mean of the child group (8.8 yr as compared to 7.9 yr, respectively), consistent with the hypothesis that performance improves over this age range. A regression analysis failed to identify a significant age effect, however ($F_{1,13}=1.79$, $p=0.20$). A post-hoc power analysis indicates that if these data are representative of the population, it would be necessary to run approximately 55 listeners to obtain a significant correlation using a one-tailed criterion of $\alpha=0.05$, with a power of 0.8 (Cohen, 1988). A line fit to the resulting data would be quite poor at describing the time course of development in detail, however, with many more listeners necessary to reliably estimate the age at which the average performance becomes adult-like. These observations are qualitatively consistent with the individual differences previously noted in intensity discrimination in children (e.g., Jensen and Neff, 1993). It is difficult to assess whether the *tone-steady* thresholds reported here are consistent with the literature on intensity discrimination in school aged children given the variability both across studies and within the current dataset, but the present results are more consistent with the suggestion that development extends out to 10 yr of age (e.g., Fior and Bolzonello, 1987) than the suggestion that intensity discrimination may be adult-like in most children by age 5 yr (e.g., Jensen and Neff, 1993).

IV. SUMMARY AND CONCLUSIONS

In Experiment 1 thresholds for a tone-in-noise detection task were different for children and adults. Not only were thresholds elevated for children, but adults benefited more from stimulus features of low-fluctuation noise than children. Results of reduced-cue conditions revealed that both children and adults performed similarly with just level cues and with tone-in-noise conditions, which also provided level-invariant cues. Most adults could make use of level-invariant cues, but most of the children were at a ceiling on these tasks. While this result leaves open the possibility that some adults could have made use of multiple cues in a simple tone in noise task, performance with just the level cue was not significantly different from the case where all cues were present, suggesting that level-invariant cues likely contributed little to performance. There was an interaction between the stimulus type and the listener group for both conditions, where the level served as a cue to the presence of a signal; that is, adults performed better in the low-fluctuation than the Gaussian noise, while children showed no effect of stimulus type. This result led to the somewhat counterintuitive hypothesis tested in Experiment 2 that children are less susceptible to the effects of external stimulus variability than are adults.

Experiment 2 estimated intensity discrimination for a tonal standard and a Gaussian noise standard, with and without an interval-by-interval level rove applied to that standard. As predicted, data from child listeners showed less evidence of threshold elevation in response to stimulus variability. The general form of the data was fitted with a simple model of performance as a function of internal and external noise. The estimate of internal noise for children was a factor of 2.5 times greater than that estimated for adults, and the effect of this greater internal noise was to temper the effects of external noise on the performance of child listeners.

While the definition of internal noise has traditionally been quite broad, encompassing variability due to transduction noise, memory, criteria, and attention—in short, any variability not accounted for by external variability—the proposal tested here is more specific, namely that the neural representation of intensity is more variable in children than adults. The conclusion that poorer performance and greater internal noise, either described qualitatively (as in Exp 1) or modeled quantitatively (as in Exp 2), can account for an age effect is important to the interpretation of the result because it does not require that we think of development in terms of the maturation of multiple specialized processing abilities. Rather, a refinement in the auditory ability may appear to emerge at different ages simply by virtue of their sensitivity to one or a small set of variables underlying internal noise. More work is needed to assess the extent to which child listeners adhere to the assumptions underlying the simple signal detection model fitted here.

America in Nashville, TN [Buss, Hall, and Grose (2003) "Developmental MLD effects," J. Acoust. Soc. Am., 113, 2207]. We would like to thank Walt Jesteadt for helpful comments on this work.

APPENDIX A: ESTIMATING THRESHOLD BASED ON INTERNAL AND EXTERNAL NOISE

Thresholds in the conditions other than the *tone-steady* condition are substantially affected by stimulus variability, quantified as σ_e in Eq. (1). Implicit in this equation is the assumption that the noise is normally distributed around the mean, such that σ_e represents the standard deviation of a Gaussian distribution. The noise associated with stimulus fluctuation in the Gaussian noise stimuli was approximately normally distributed, with a slight skew (third moment) of −0.68, when represented in dB. The noise introduced by the rove, however, was clearly not normal, being based on a uniform distribution applied to the standard.

Rather than approximating external noise as normally distributed and predicting thresholds with Eq. (1), a MATLAB routine was written to generate cue distributions associated with the stimuli used here. At each of a range of signal levels, 5000 samples of signal present and 5000 samples of no-signal stimuli were generated using the same procedures as employed in Experiment 2. Independent samples of Gaussian-distributed internal noise were then added to each estimate of signal level in each array, based on estimates of σ_i computed above. Pairs of signal-present and no-signal samples were subtracted, for a total of 5000 values of cue difference. The distribution of these differences was then used to estimate d', according to the formula

$$d' = \sqrt{2}\Delta/\sigma_{\text{diff}}, \tag{A1}$$

where Δ is the signal level associated with the signal-present array and σ_{diff} is the standard deviation of the cue difference array. The scalar ($\sqrt{2}$) is introduced because the standard deviation of the difference between two distributions is $\sqrt{2}$ times the standard deviation of each contributing distribution, assuming they are of equal variance. In this way the d' associated with each signal level was computed. A total of 5 signal levels were used, spaced at 0.5 dB intervals and chosen adaptively to bracket the final estimate. The relationship between d' and signal level was well characterized in terms of a least-squares line fit, and this fitted line was used to estimate the signal level associated with $d' = 1.61$.

This process was repeated 8 times: 4 levels of external variability (*tone-steady, tone-roved, Gnoise-steady, and Gnoise-roved*) × two groups (child and adult, with $\sigma_i = 2.45$ and $\sigma_i = 0.99$, respectively). To give an indication of the reliability of these procedures, threshold estimates in the *tone-steady* condition computed using this procedure were within 0.1 dB of those computed analytically. Threshold predictions are indicated with dark lines in Fig. 3.

[1] Due to time constraints, one child listener did not repeat the *all-cues* condition, so his data are not represented in this analysis.
[2] Buss *et al.* (2003) ran 13 child listeners, two of whom also participated in Experiment 1.

[3] Elimination of these data points did not change the general conclusions of the experiment. Repeating the statistics with outlier data included produced the same pattern of significance as reported in Sec. III B.
[4] The primary consideration in selection of units for intensity discrimination, apart from ease of comparison with other data, was compliance with assumptions regarding the variance across the dataset. For the ANOVA, Box's Test of equality of covariance was not significant ($p=0.10$) for data represented in dB SPL. Units of ΔL were less consistent with assumptions of normal and uniform variance: in that case Box's Test was significant ($p<0.05$), indicating that the covariance matrices were not equal across groups. This result motivated the use of dB SPL units in the statistical tests reported here.

Allen, P., Jones, R., and Slaney, P. (**1998**). "The role of level, spectral, and temporal cues in children's detection of masked signals," J. Acoust. Soc. Am. **104**, 2997–3005.

Allen, P., and Nelles, J. (**1996**). "Development of auditory information integration abilities," J. Acoust. Soc. Am. **100**, 1043–1051.

Allen, P., and Wightman, F. (**1994**). "Psychometric functions for children's detection of tones in noise," J. Speech Hear. Res. **37**, 205–215.

Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

ANSI (**1996**). (S3-1996), "Specification for Audiometers," American National Standards Institute, New York.

Bos, C. E., and deBoer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Buss, E., Hall III, J. W., and Grose, J. H. (**2003**). "Developmental MLD effects," J. Acoust. Soc. Am. **113**, 2207.

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Lawrence Erlbaum, Hillsdale, NJ).

Eddins, D. A., and Barber, L. E. (**1998**). "The influence of stimulus envelope and fine structure on the binaural masking level difference," J. Acoust. Soc. Am. **103**, 2578–2589.

Eddins, D. A. (**2001**). "Monaural masking release in random-phase and low-noise noise," J. Acoust. Soc. Am. **109**, 1538–1549.

Elliot, L. L. (**1979**). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," J. Acoust. Soc. Am. **66**, 651–653.

Fior, R. (**1972**). "Physiological maturation of auditory function between 3 and 13 years of age," Audiology **11**, 317–321.

Fior, R., and Bolzonello, P. (**1987**). "An investigation on the maturation of hearing abilities in children," Ear Hear. **8**, 347–349.

Hall III, J. W., and Grose, J. H. (**1994**). "Development of temporal resolution in children as measured by the temporal modulation transfer function," J. Acoust. Soc. Am. **96**, 150–154.

Hall III, J. W., Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.

Hall III, J. W., Grose, J. H., and Dev, M. B. (**1997**). "Auditory development in complex tasks of comodulation masking release," J. Speech Lang. Hear. Res. **40**, 946–954.

Hall III, J. W., Grose, J. H., and Hartmann, W. M. (**1998**). "The masking-level difference in low-noise noise," J. Acoust. Soc. Am. **103**, 2573–2577.

Hartmann, W. M., and Pumplin, J. (**1988**) "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.

Jensen, J. K., and Neff, D. L. (**1993**). "Development of basic auditory discrimination in preschool children," Psychol. Sci. **4**, 104–107.

Jesteadt, W., Nizami, L., and Schairer, K. S. (**2003**). "A measure of internal noise based on sample discrimination," J. Acoust. Soc. Am. **114**, 2147–2157.

Kidd, G. Jr., Mason, C. R., Brantley, M. A., and Owen, G. A. (**1989**). "Roving-level tone-in-noise detection," J. Acoust. Soc. Am. **86**, 1310–1317.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., and Oxenham, A. J. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acta Acust. **83**, 659–669.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics." J. Acoust. Soc. Am. **49**, 467–477.

Maxon, A. B., and Hochberg, I. (**1982**). "Development of psychoacoustic behavior: Sensitivity and discrimination," Ear Hear. **3**, 301–308.

Nábelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Buss *et al.*: Development and internal noise   2787

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.

Olsho, L. W., Koch, E. G., Halpin, C. F., and Spetner, N. B. (**1988**). "Pure-tone sensitivity of human infants," J. Acoust. Soc. Am. **84**, 1316–1324.

Rice, S. O. (**1953**). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York), pp. 133–294.

Richards, V. M. (**2001**). "Varying feedback to evaluate detection strategies; the detection of a tone added to noise," J. Assoc. Res. Otolaryngol. **3**, 209–221.

Richards, V. M. (**1992**). "The detectability of a tone added to narrow bands of equal-energy noise," J. Acoust. Soc. Am. **91**, 3424–3435.

Richards, V. M., and Nekrich, R. D. (**1993**). "The incorporation of level and level-invariant cues for the detection of a tone added to noise," J. Acoust. Soc. Am. **94**, 2560–2574.

Schneider, B. A., Morrongiello, B. A., and Trehub, S. E. (**1990**). "Size of critical band in infants, children and adults," J. Exp. Psychol. Hum. Per-cept. Perform. **16**, 642–652.

Schneider, B. A., and Trehub, S. E. (**1992**). "Sources of developmental change in auditory sensitivity," in *Developmental Psychoacoustics*, edited by L. A. Werner, and E. W. Rubel (American Psychological Association, Washington, DC), pp. 3–46.

Schneider, B. A., Trehub, S. E., Morrongiollo, B. A., and Thorpe, L. A. (**1989**). "Developmental changes in masked thresholds," J. Acoust. Soc. Am. **86**, 1733–1742.

Spiegel, M. G., and Green, D. M. (**1981**). "Two procedures for estimating internal noise," J. Acoust. Soc. Am. **70**, 69–73.

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.

Willihnganz, M. S., Stellmach, M. A., and Lutfi, R. A. (**1997**). "Spectral weights in level discrimination by preschool children: Synthetic listening conditions," J. Acoust. Soc. Am. **101**, 2803–2810.

Zhang, C., and Zeng, F.-G. (**1997**). "Loudness of dynamic stimuli in acoustic and electric hearing," J. Acoust. Soc. Am. **102**, 2925–2934.

Frequency and amplitude discrimination along the kinesthetic-cutaneous continuum in the presence of masking stimuli[a)]

Ali Israr[b)] and Hong Z. Tan
Haptic Interface Research Laboratory, Purdue University, 465 Northwestern Avenue, West Lafayette, Indiana 47907

Charlotte M. Reed
Research Laboratory of Electronics, Room 36-751, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Frequency and amplitude discrimination thresholds along the kinesthetic to cutaneous continuum were evaluated on the left index fingerpad using a multifinger tactual display. Target stimuli were presented either in isolation (no-masker condition) or in the presence of masking stimuli (one- or two-masker conditions). Six reference target signals in the frequency range 2–300 Hz (two each from low-, medium-, and high-frequency regions) and at an amplitude of either 20 or 35 dB sensation levels (SL) were used. In the no-masker condition, the range of frequency Weber fraction was 0.13–0.38 and 0.14–0.28, and the range of amplitude discrimination threshold was 1.82–2.98 dB and 1.65–2.71 dB, at 20 and 35 dB SL, respectively. In the masking conditions, average frequency Weber fractions rose to 0.60 and 0.46, and average amplitude thresholds rose to 3.63 and 3.72 dB, at 20 and 35 dB SL, respectively. In general, thresholds were largest in the two-masker condition and lowest in the no-masker condition. Although the frequency and amplitude thresholds generally increased in the presence of masking stimuli, there was some indication of channel independence for low- and high-frequency target stimuli. The implications of the results for tactual communication of speech are discussed. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354022]

I. INTRODUCTION

The present study is motivated by issues related to the development of artificial aids designed for the communication of acoustic signals through the sense of touch for persons with profound hearing impairment. A multifinger tactual display capable of simultaneously stimulating the thumb, index, and middle fingers, the TACTUATOR, has been developed recently to broaden the spectrum of tactual stimulation from the tactile/cutaneous region (small-amplitude high-frequency vibrations) that is commonly used by most artificial tactile aids to the kinesthetic region (large-amplitude low-frequency movements) (Tan and Rabinowitz, 1996). The idea of presenting movement signals in addition to vibrations in a tactual communication aid was inspired by a means of natural tactual speech communication called the Tadoma method where a deaf-and-blind receiver places one hand or both hands on the face of a talker and monitors the tactual signals (e.g., laryngeal vibration, mouth opening, lip movements, air flow at the lips, etc.) associated with the articulatory process (Reed *et al.*, 1985). One hypothesis for the success of the Tadoma method is that Tadoma users have access to a rich set of signals including both cutaneous and kinesthetic stimulation. The TACTUATOR is uniquely suited for evaluating such a hypothesis because unlike other tactual displays (such as a minishaker) that can generate signals at very low frequencies but can only do so at near-threshold intensity levels, the TACTUATOR is capable of delivering slow motions at a peak-to-peak displacement of up to 25 mm.

Previous research conducted with the TACTUATOR shows high-information transmission capability with stimulus sets composed of kinesthetic movements, cutaneous vibrations, signals with intermediate frequencies, and their combinations. For example, a static information transfer (IT) of 6.5 bits and a dynamic IT rate of 12 bits/sec were achieved in one study (Tan *et al.*, 1999) and an IT rate of 21.9 bits/sec was reported in another (Tan *et al.*, 2003). These static IT and dynamic IT rates are among the highest reported for any artificial tactual display and are comparable to those of the natural Tadoma method used by deaf-and-blind individuals for speech communication (Reed and Durlach, 1998).

A subsequent study using the TACTUATOR evaluated temporal onset-order discrimination with a 50-Hz sinusoidal signal delivered to the thumb and a 250-Hz sinusoidal vibration to the index finger (Yuan, Reed, and Durlach, 2005a). The average temporal onset-order threshold was found to be 34 msec. The same authors then used amplitude envelope information from two spectral regions of speech to transmit consonant voicing information through the TACTUATOR (Yuan, Reed, and Durlach, 2005b). The amplitude envelope of the high-frequency band was used to modulate a 250-Hz vibration at the index finger and the amplitude envelope of

---

the low-frequency band was used to modulate a 50-Hz vibration at the thumb. Results indicated that noise-masked normal-hearing subjects were highly successful in discriminating voiced from unvoiced sounds at the segmental level and that the display provided a substantial benefit to lip-reading alone in closed-set consonant identification tasks. Results of sentence recognition tasks, however, revealed that the tactual cues were not beneficial for lip reading. This lack of benefit may have reflected the user's inability to integrate tactual and visual cues and therefore may have been the result of insufficient training.

Recently, a 2-DOF (degree of freedom) controller has been developed for the TACTUATOR that preserves the relative intensities of the spectral components of input waveforms in terms of perceived intensity (Israr, Meckl, and Tan, 2004). Consider the case where the input signal to the TACTUATOR contains two spectral components at 50 and 250 Hz at equal voltage levels. The 2-DOF controller compensates for both the frequency response of the TACTUATOR hardware and the difference in human detection threshold at the respective frequencies such that the two signal components are perceived at roughly the same sensation levels (SL). With this controller, a low-pass filtered broadband acoustical speech signal can be directly applied as the input to the TACTUATOR to transmit, for example, pitch information. The use of such an encoding scheme for speech obviously requires that the user be able to discern variations in frequency and/or amplitude in the presence of other masking signals occurring before, after, or simultaneously with the target. In order for speech signals to be transmitted successfully through the TACTUATOR, it is important to characterize frequency and amplitude discrimination thresholds associated with the use of the TACTUATOR in the presence of masking stimuli delivered to the same or neighboring digits of the hand.

Previous studies have examined frequency and amplitude discrimination thresholds through the tactual sense. Frequency discrimination thresholds, typically expressed in the form of Weber fractions (WFs), computed as $(\Delta F)_0 / F_{ref}$ where $(\Delta F)_0$ is the frequency discrimination threshold in hertz and F_{ref} the reference frequency, have been found to be in the range 0.02 to 0.72 on fingertips, forearm, thenar eminence, and sternum using sinusoidal or pulselike waveforms over a frequency range of 1 to 512 Hz and an amplitude range of 14 to 35 dB SL (Franzen and Nordmark, 1975; Globe and Hollins, 1994; Goff, 1967; Horch, 1991; Knudsen, 1928; LaMotte and Mountcastle, 1975; Rinker, Craig, and Bernstein, 1998; Rothenberg et al., 1977). Variability in the size of frequency WFs across studies has been attributed to differences in experimental conditions, methodology, and stimulus artifacts (Bensmaia et al., 2005; Hnath-Chisolm and Medwetsky, 1988; Rothenberg et al. 1977; Verrillo and Gescheider, 1992).

Amplitude discrimination thresholds are typically expressed in decibels in amplitude change. Previous studies of just-noticeable-differences (JNDs) in intensity for tactile stimulation have employed a variety of experimental methods (continuous versus gated methods, pulses versus pure sinusoids, forced-choice method versus method of adjustment, method of limits versus adaptive procedures) using

stimuli that varied in frequency, amplitude, body site, and contact area. Amplitude-discrimination thresholds in $(\Delta A)_0$ have been reported in the range of roughly 0.4 to 6 dB; for example, 0.8 dB (Geldard, 1957), 1.5–2.5 dB (Craig, 1972), 1.5–6 dB (Craig, 1974), 0.4–2 dB (Fucci, Small, and Petrosino, 1982), 0.5–4 dB (Gescheider et al., 1990), 0.8–5 dB (Gescheider, Zwislocki, and Rasmussen, 1996), 0.8–3 dB (Globe and Hollins, 1993), and 0.8–1.5 dB (Rinker et al., 1998). Despite differences in experimental conditions, it is generally agreed that amplitude-discrimination thresholds tend to decrease with an increase in stimulus intensity when the intensity is close to detection threshold, and then stay constant when stimulus intensity is above 14 to 20 dB SL (see, for example, Craig, 1972; Craig, 1974; Gescheider et al., 1990).

Despite all existing data, there have been relatively few studies concerned with the masking properties of kinesthetic signals on vibrations and vice versa, presumably because, until very recently, no displays existed that could deliver signals over a wide range of frequencies that spanned the kinesthetic-cutaneous continuum.[1] The TACTUATOR provides a unique opportunity for studying frequency and amplitude discrimination thresholds using both kinesthetic and cutaneous signals in the presence of masking stimuli. In an earlier study, experiments were conducted to examine the temporal masking properties for recognition of tactual stimulation patterns using forward masking, backward masking, and sandwiched masking (in which the target was presented between two maskers) paradigms (Tan et al., 2003). In the present study, frequency and amplitude discrimination experiments were performed in the presence of masking stimuli that occurred *simultaneously* with the target. The results from the present and previous studies will shed light on the design of multidimensional tactual signals for speech communication.

II. METHODS

A. Apparatus

The experimental apparatus (the TACTUATORII[2]) consists of three independent, point-contact (with a contact area of approximately 70 mm²), one-degree-of-freedom actuators interfaced individually with the fingerpads of the thumb, the index finger, and the middle finger [Fig. 1(a)]. The motion trajectory for the thumb is perpendicular to that of the index and middle fingers, thereby maintaining an approximately natural hand configuration with the wrist resting in its neutral position. The range of motion for each digit is about 25 mm. All motions begin and end with each of the three digits at the middle of its respective range of motion. Each digit can thus be moved either outward (extension) or inward (flexion).

Each actuator utilizes a disk-drive head-positioning motor augmented with angular position feedback from a precision rotary variable differential transformer that has a response bandwidth (−3 dB) of 1 kHz and a theoretically unlimited resolution[3] due to electromagnetic coupling [Fig. 1(b)]. The overall system performance is well suited for the present study for several reasons. First, each movement channel has a *continuous* frequency response from dc to 300 Hz (the disk-drive motor ceases to move beyond about
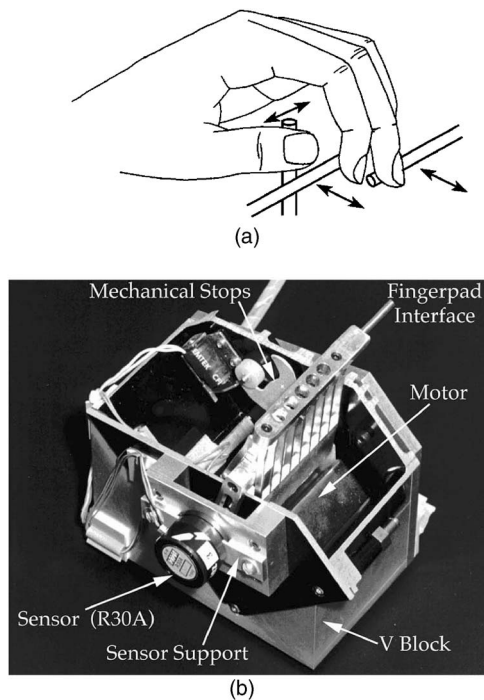
FIG. 1. The TACTUATOR. (a) Schematic drawing illustrating finger placement on the TACTUATOR and the motion trajectories of the thumb, index, and middle fingers. (b) Photograph of one of the three motor assemblies with its components labeled. The participant rests the fingerpad on the rod labeled "finger interface" in (b).

400 Hz). Therefore, stimulation can be delivered in the kinesthetic (i.e., low-frequency gross motion) and cutaneous (i.e., high-frequency vibration) ranges as well as in the mid-frequency range. Stimuli with any desired spectral components can be realized within the frequency range of dc to 300 Hz. Second, across the frequency range of dc to 300 Hz, amplitude of 0 dB SL to at least 47 dB SL can be achieved at each frequency.[4] This stimulus range is well matched to the dynamic range of tactual perception in that stimulation levels exceeding 50–55 dB SL tend to induce discomfort and fatigue (Verrillo and Gescheider, 1992). Third, measurements with single and multiple frequency inputs at various amplitude levels indicate that each channel is highly linear, harmonic distortion is low, and interchannel crosstalk is small. These characteristics permit high-fidelity delivery of waveforms of arbitrary frequency content and stimulation level (e.g., a 20 dB SL high-frequency vibration superimposed on a 35 dB SL low-frequency motion) simultaneously to all fingers. Fourth, "loading" a movement channel (i.e., resting a finger lightly on the actuator's moving bar) does not significantly alter the intended stimuli. Selected measurements indicate that loading reduces the magnitude of stimulation by an average of 1.5 dB at 2 Hz, 2.7 dB at 20 Hz, and 0.1 dB at 200 Hz. These minor alterations due to loading do not pose a significant problem for the experiments reported here because the stimuli are generally strong (i.e., at least 20 dB SL).

Real-time positional control is provided by a digital 2-DOF controller implemented on a floating-point digital signal processing (DSP) system with 16-bit analog-to-digital and digital-to-analog converters. The controller shapes the frequency response of the closed-loop TACTUATORII assembly to that of the human detection-threshold function, thus preserving the relative amplitude of spectral components in terms of their perceived intensity by human observers. The current system can therefore be driven with a broadband signal (up to 300 Hz) while preserving the relative intensity of different spectral components in terms of the relative sensation levels delivered by the TACTUATORII (Israr et al., 2004). The nominal threshold function values for six target frequencies are shown in Table I (see last column). These thresholds are averages of measurements obtained by Israr et al. (2004) on three experienced observers.

B. Participants

Four participants (2 females and 2 males; age 21 to 28 years old; average age 24 years old) took part in the experiments. S1 was highly experienced with the TACTUATORII device. S2 and S4 had not felt the device prior to the present study. S1 and S3 were research staff and had previous experience in other haptic perception experiments. S2 and S4, who had no prior experience with psychophysical experiments, were paid on an hourly basis. None of the participants reported any known tactual impairments of their hands.

C. Stimuli

The present study employed a stimulus set applied only to the left index finger. The duration of each stimulus was fixed at 250 msec throughout the experiments. The independent stimulus attributes were frequency and amplitude of sinusoidal signals. The entire frequency range of the TACTUATORII (0–300 Hz) was divided into three regions (low, medium, and high) as in previous studies with the TACTUATOR device (Tan et al., 1999; Tan et al., 2003). Two target frequencies were tested in each frequency region at two amplitudes of 20 and 35 dB SL (sensation level; decibels above nominal detection threshold at the same frequency). A 250-ms Hanning window (with 25-ms rise and fall time) was used so that each stimulus began and ended in the middle of the finger's range of motion. All stimuli started in the finger extension direction. A complete list of the test stimuli is provided in Table I. Also shown are the nominal detection threshold levels measured from three participants (one of whom was S1) in our previous study (Israr et al., 2004) from which signal amplitudes in decibels SL are derived for the present study. The sensation levels reported in Table I are

TABLE I. Listing of stimuli used in the frequency and amplitude discrimination experiments.

Frequency Region	Frequency (Hz)	Amplitude (dB SL)	Nominal Detection Threshold (dB re 1 μm peak)
Low	2	20 & 35	38
	4	20 & 35	34
Medium	15	20 & 35	25
	30	20 & 35	16
High	80	20 & 35	−8
	200	20 & 35	−18

nominal and representative, but not individualized for each participant (see the Appendix).

Different perceptual attributes are associated with the stimuli within each region. The 2- and 4-Hz low-frequency signals are perceived as movements that extend the index finger and bring it back to its resting position. The 15- and 30-Hz mid-frequency signals give rise to a mixed flutter/rough sensation. The 80- and 200-Hz high-frequency signals are perceived as smooth and penetrating vibrations. Within each of the three frequency regions, the two test stimuli can be easily discriminated in isolation.

1. Stimuli for Frequency Discrimination

For frequency discrimination, the participant was presented with either the *reference* frequency F_{ref}, or the *test* frequency $F_{\text{ref}}+\Delta F$, where ΔF was a variable frequency increment. Each F_{ref} was presented with either $A_1=20$ dB SL or $A_2=35$ dB SL. During an experimental run, the same amplitude (in sensation level) was used for F_{ref} and $F_{\text{ref}}+\Delta F$. The *target* stimuli (F_{ref} or $F_{\text{ref}}+\Delta F$) were presented either in isolation (no-masker condition $C1$) or in the presence of a *masker* that occurred simultaneously with the target (one-masker conditions $C2$ and $C3$, and two-masker condition $C4$). $C1$ served as a baseline condition against which the effect of masking stimuli could be measured. For $C2$, a single masker was selected from the lower one of the two frequency regions of which F_{ref} was not a member. For $C3$, a single masker was selected from the other (i.e., higher) frequency region that did not include F_{ref}. For example, if $F_{\text{ref}} =2$ Hz, then the masker in $C2$ was selected from the mid-frequency region, and the masker in $C3$ from the high-frequency region. For $C4$, two maskers were selected, one from each of the two frequency regions that did not include the target F_{ref}. In any of the masking conditions ($C2–C4$), the frequency of the masker was selected randomly from ten values linearly spaced between the two frequency values listed for that frequency region as shown in Table I. For example, the frequency of a masker in the mid-frequency region assumed values in the range of 15 to 30 Hz. The amplitude of the masker was also selected at random on each trial. In order that the intensity of the masker did not overwhelm the perception of the target signal, the masker amplitude for a target presented at $A_1=20$ dB SL was randomly selected from ten levels linearly spaced between 10 and 20 dB SL. The masker amplitude for a target presented at $A_2=35$ dB SL was randomly selected from ten levels between 20 and 35 dB SL. For any of the masking conditions, the masker frequency and amplitude were randomly selected on each presentation.

2. Stimuli for Amplitude Discrimination

The stimuli for amplitude discrimination were very similar to those used in frequency discrimination. The participant was presented with either the *reference* amplitude A_{ref}, or the *test* amplitude $A_{\text{ref}}+\Delta A$, where ΔA was a variable amplitude increment. During an experimental run, the same frequency was used for A_{ref} and $A_{\text{ref}}+\Delta A$. Two values of A_{ref} were tested: 20 and 35 dB SL. Six values of frequency were employed in the amplitude-discrimination tests consisting of two values from each of the three frequency regions as listed in Table I. As described above for the frequency discrimination experiment, the target stimuli were presented either in isolation ($C1$) or in the presence of masking stimuli ($C2–C4$). The amplitude-discrimination experiments also employed randomized values of masker frequency and amplitude on each presentation using the procedure described above for the frequency-discrimination experiments.

D. Procedures

The same experimental procedure was used in both the frequency and amplitude discrimination experiments. A three-interval forced-choice (3IFC) paradigm with a one-up three-down adaptive procedure (Leek, 2001; Levitt, 1971) was used. On each trial, the participant received three stimulus presentations. One randomly selected interval contained the test stimulus ($F_{\text{ref}}+\Delta F$ or $A_{\text{ref}}+\Delta A$) while the remaining two intervals contained the reference stimulus (F_{ref} or A_{ref}). During the masking conditions ($C2–C4$), the frequency and amplitude of the masker were randomly selected for each of the three intervals. The participant's task was to indicate the "odd" interval that contained $F_{\text{ref}}+\Delta F$ or $A_{\text{ref}}+\Delta A$ by pressing "1", "2", or "3" on the keyboard. The initial value of ΔF or ΔA was selected to be sufficiently above the expected discrimination threshold. The value of ΔF (or ΔA) changed by 2/3 octave (or 4 dB) initially (for faster convergence), and then by 1/6 octave (or 1 dB) after the first three reversals (for better resolution of threshold estimate). The experimental run terminated after 12 reversals at the smaller step size, which typically took 60–80 trials. After each experimental run, a plot of ΔF or ΔA as a function of trial number was generated. If it appeared that the stimulus frequency (or amplitude) did not converge, the participant was asked to repeat the same run again. Each run lasted about 4–6 min. Participants typically completed four experimental runs (one reference stimulus with all four conditions $C1–C4$) in one session.

For frequency discrimination, there were 12 possible reference signals (3 frequency regions × 2 frequencies per region × 2 amplitudes per frequency). The order of the reference signals was randomized for each participant. Given each reference signal, however, the four conditions $C1–C4$ were always tested in the same order as follows. The no-masker condition $C1$ was tested first, followed by either one of the one-masker conditions $C2$ or $C3$ (randomly chosen for each reference stimulus), then the remaining $C2$ or $C3$ that had not been tested yet, and finally the two-masker condition $C4$. Participants were required to complete all four conditions associated with one reference stimulus in one session. Assuming that a typical adaptive run terminated after 70 trials, each participant completed a total of 3360 trials (12 references × 4 conditions per reference × 70 trials per condition) for the frequency discrimination experiment. For amplitude discrimination, there were also 12 possible reference signals. Again, the order with which the reference signals were presented was randomized, and participants always went through $C1–C4$ in the same order for each reference
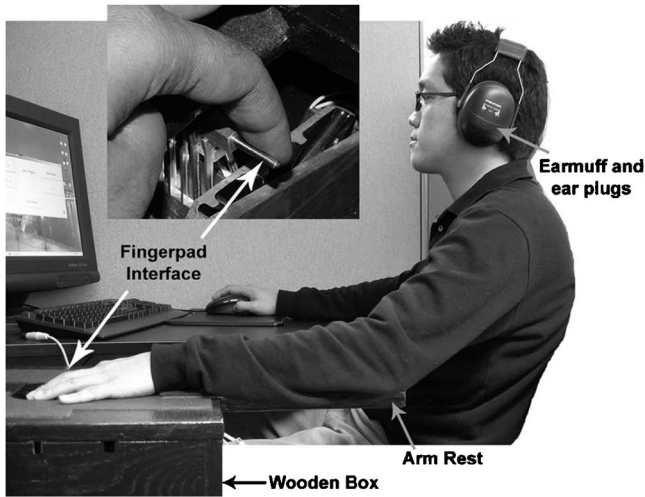
FIG. 2. Experimental setup. The apparatus was placed to the left of the participant's torso in a wooden box. The participant sat in front of the computer screen and placed the left index finger on the finger rest of the TACTUATORII as shown in the inset. The left elbow was supported by the arm rest. The participant wore earmuffs and ear plugs.

signal as in the frequency discrimination experiment.[5] Data were collected on a total of 48 conditions in both the frequency and the amplitude discrimination experiments. Thus, each participant completed a total of 96 experimental runs which required roughly 6,720 trials.

During the experiments, the TACTUATORII was placed to the left of the participant's torso as shown in Fig. 2. It was covered by a padded wooden box that served as an armrest for the participant's left forearm. The top of the box had an opening that allowed the participant to place the left index finger on the "fingerpad interface" rod as shown in Figs. 1(b) and 2. Earmuff (Twin-Cup type, H10A, NRR 29, Peltor, Sweden) and pink noise (presented at roughly 80 dB SPL) were used to eliminate possible auditory cues. No correct-answer feedback was provided during the experiments.

Training was provided before each run. During training, the participant could click on the appropriate buttons on a computer screen to feel the reference (F_{ref} or A_{ref}) or the test stimulus ($F_{ref}+\Delta F$ or $A_{ref}+\Delta A$). For the masking conditions $C2-C4$, the participant could either feel the target stimulus (reference or test signal) in isolation or in the presence of a masker (as would be the case during the actual experiment). The participant could terminate the training whenever she/he was ready. Training typically lasted only a few minutes. Data collection followed immediately after the training.

E. Data analysis

From the six pairs of peaks and valleys of stimulus intensities (ΔF or ΔA) obtained during the last 12 reversals at the smaller step size, six estimates of thresholds were obtained by averaging each peak-valley pair. The frequency and amplitude detection thresholds, $(\Delta F)_0$ in hertz and $(\Delta A)_0$ in decibels, respectively, were then estimated by averaging the six threshold estimates. The corresponding standard errors were also obtained from the six threshold estimates. The threshold so obtained corresponds to the 79.4-percentile

point on the psychometric function (Levitt, 1971), and a d' value of 1.63 assuming no response bias (Hacker and Ratcliff, 1979).

III. RESULTS

A. Frequency discrimination thresholds

Under the no-masker baseline condition ($C1$), the frequency discrimination threshold, $(\Delta F)_0$ in hertz, increased with the reference frequency F_{ref}. The linear regression model was $(\Delta F)_0=0.13\cdot F_{ref}+2.7$ ($r^2=0.7142$) at 20 dB SL, and $(\Delta F)_0=0.15\cdot F_{ref}+0.7$ ($r^2=0.8415$) at 35 dB SL. The average Weber fractions, WF=$(\Delta F)_0/F_{ref}$, are shown for each condition (in four separate panels) in Fig. 3 for 20 dB SL and in Fig. 4 for 35 dB SL. A visual inspection of both Figs. 3 and 4 revealed large interparticipant differences (particularly for the three masking conditions $C2-C4$), confirmed by a four-way (participant, frequency, amplitude, condition) analysis-of-variance (ANOVA) ($p<0.0001$). The Scheffe *post hoc* multirange test indicated two participant groups: $S1$ and $S3$ (experienced, with lower thresholds), and $S2$ and $S4$ (inexperienced, with higher thresholds). Despite the significant differences among participants, however, all participants' data displayed similar trends along frequency, amplitude, and masking conditions. Therefore, we discuss the general trends demonstrated by the pooled data and point out individual exceptions when they are significant.

The frequency WFs from the no-masker condition [$C1$; see Figs. 3(a) and 4(a)] averaged 0.25 and 0.18 at 20 and 35 dB SL, respectively. A two-way (amplitude, frequency) ANOVA showed that both frequency and amplitude were significant ($p<0.001$). Specifically, the WFs were significantly lower at 35 dB SL than at 20 dB SL, and they generally decreased as F_{ref} increased. The interaction term was also significant ($p<0.001$) indicating mixed trends along frequency and amplitude. For low-frequency targets ($F_{ref}=2$ and 4 Hz), the WFs were significantly lower at 35 dB SL than at 20 dB SL. For targets at $F_{ref}=200$ Hz, however, the opposite trend was observed. The trends were mixed for targets at $F_{ref}=15$, 30, and 80 Hz.

When frequency WFs from all conditions ($C1-C4$; see Figs. 3 and 4) were considered, it was found that the average WF at 35 dB SL (mean = 0.39) was significantly lower than that at 20 dB SL (mean = 0.52) by a four-way (participant, frequency, amplitude, condition) ANOVA ($p<0.001$). As expected, the WFs for the no-masker condition ($C1$) were no greater than those for the one-masker conditions ($C2$ or $C3$), which in turn were no greater than those for the two-masker condition ($C4$), as confirmed by the four-way ANOVA ($p<0.001$). The average WFs decreased significantly in the order $C4$, $C2$, $C3$, and $C1$ (average WFs were 0.70, 0.49, 0.40, and 0.22, respectively). All the interaction terms of the four-way ANOVA were significant ($p<0.001$) implying mixed trends along frequency, amplitude, and masking conditions.

To visualize the masking effects ($C2-C4$), the average frequency WFs for $C1$ were subtracted from the corresponding WFs for $C2-C4$. The differences are shown in Fig. 5 for stimuli presented at (a) 20 dB SL and (b) 35 dB SL, respec-
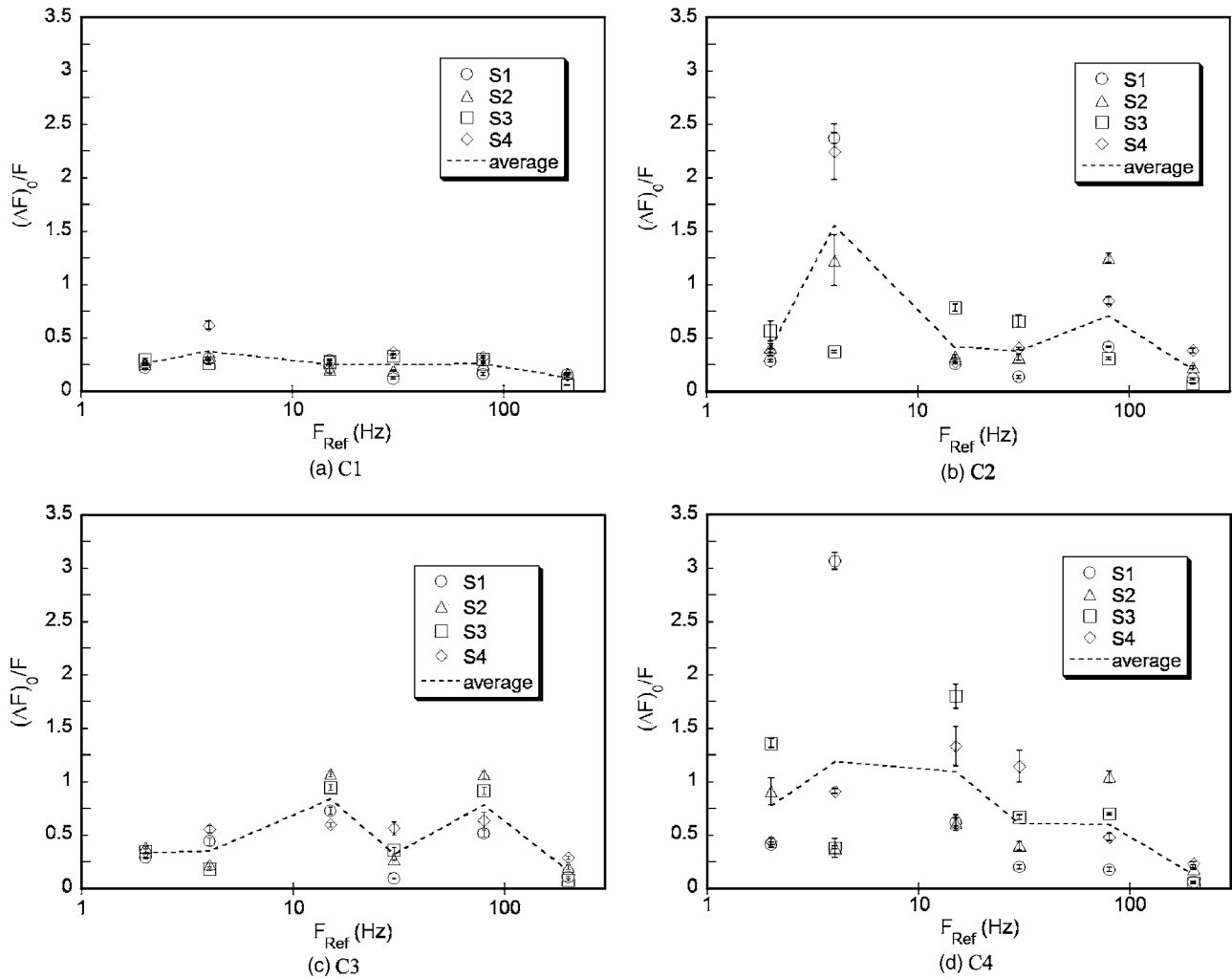
FIG. 3. Frequency discrimination Weber fractions $(\Delta F)_0 / F_{ref}$ for the four participants as a function of reference frequency F_{ref} at 20 dB SL. Panel (a) shows the thresholds in the no-masker condition C1. Panels (b), (c), and (d) show thresholds for the one- and two-masker conditions C2, C3, and C4, respectively. Shown are individual participant's data (open symbols), the average (dashed lines), and standard errors.

tively. An asterisk indicates that the corresponding data column is significantly different from 0 ($p < 0.05$). A four-way ANOVA analysis (participant, frequency, amplitude, condition) applied to the data in Fig. 5 showed that there were more masking effects at 20 dB SL than at 35 dB SL ($p < 0.001$). All interaction terms were again significant ($p < 0.001$). The WFs for low-frequency targets ($F_{ref} = 2$ and 4 Hz) increased in the presence of mid-frequency maskers ($C2-C1$), but the WFs for mid-frequency targets ($F_{ref} = 15$ and 30 Hz) were generally not affected by low-frequency maskers ($C2-C1$). The WFs for low-frequency targets were not affected by high-frequency maskers ($C3-C1$), yet the WFs for high-frequency targets ($F_{ref} = 80$ and 200 Hz) were affected by low-frequency maskers ($C2-C1$). The interference between mid-frequency and high-frequency targets/maskers appeared to depend on both frequency and amplitude.

B. Amplitude discrimination thresholds

The average amplitude thresholds $(\Delta A)_0$ in dB are plotted against reference frequency (F_{ref}) for all four masking conditions ($C1-C4$) in Fig. 6 for 20 dB SL and Fig. 7 for 35 dB SL. Each of the four panels presents amplitude thresholds for one of the four conditions. Similar to the results of frequency discrimination experiments, interparticipant variability was observed in these results. A four-way ANOVA (participant, frequency, amplitude, condition) confirmed a significant interparticipant difference ($p < 0.0001$). A Scheffe *post hoc* multirange test indicated that lower thresholds were obtained for participants $S1$ and $S3$ (experienced), and higher thresholds were obtained for participants $S2$ and $S4$ (inexperienced). Despite the significant differences among participants, however, all participants displayed similar trends along frequency, amplitude, and background conditions. Therefore, we discuss the general trends demonstrated by the pooled data and point out individual exceptions when they are significant.

The amplitude discrimination thresholds in the no-masker condition [$C1$; see Figs. 6(a) and 7(a)] averaged 2.5 dB and 2.17 dB at 20 and 35 dB SL, respectively. A two-way ANOVA (frequency, amplitude) of the data showed that both frequency and amplitude were significant (p
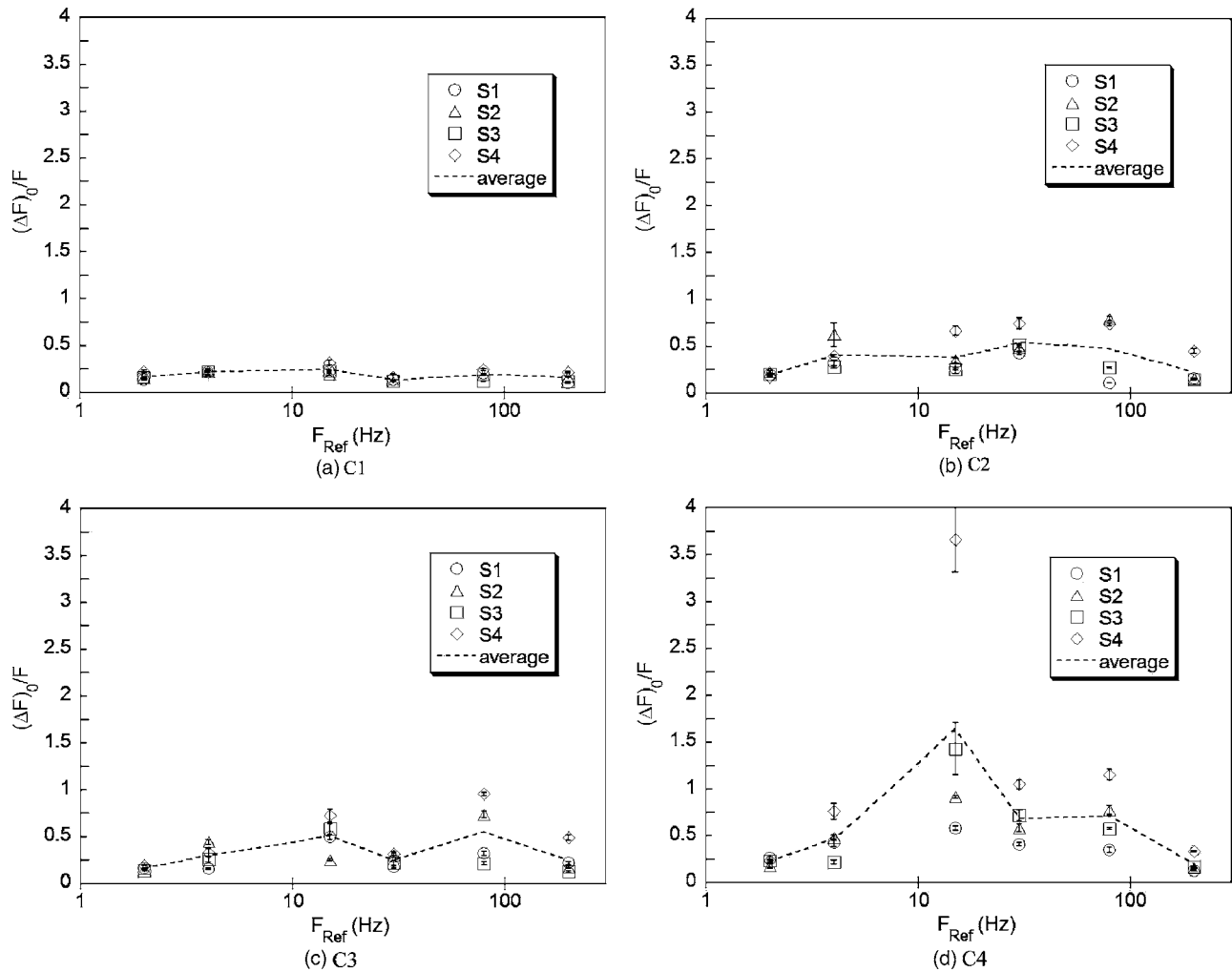
FIG. 4. Frequency discrimination Weber fractions $(\Delta F)_0/F_{ref}$ for the four participants as a function of reference frequency F_{ref} at 35 dB SL. Panel (a) shows the thresholds in the no-masker condition C1. Panels (b), (c), and (d) show thresholds for the one- and two-masker conditions C2, C3, and C4, respectively. Shown are individual participant's data (open symbols), the average (dashed lines), and standard errors.

<0.001). The interaction term was also significant (p <0.001) indicating mixed trends along frequency and amplitude. Unlike frequency discrimination data that exhibited a scattered pattern, an orderly transition of amplitude discrimination thresholds from low- to high-frequency regions was observed at both amplitude levels [see Figs. 6(a) and 7(a)]. A Scheffe multirange test at the two amplitudes (20 and 35 dB SL) separated the target frequencies into two groups. The thresholds for the low-to-mid-frequency group (F_{ref}=2, 4, and 15 Hz) were found to be significantly higher than those for the mid-to-high-frequency group (F_{ref}=30, 80, and 200 Hz) at 20 dB SL, and the opposite was true at 35 dB SL. The analysis also showed that the thresholds for the low-to-mid-frequency group were significantly lower at 35 dB SL than at 20 dB SL, but the thresholds for the mid-to-high-frequency group were similar at the two amplitude levels.

Unlike the results obtained in the frequency discrimination experiments, the average amplitude discrimination thresholds (see Figs. 6 and 7) were not significantly different at the two amplitude levels, as indicated by a four-way (participant, frequency, amplitude, condition) ANOVA (p >0.05). Similar to the frequency discrimination thresholds,

masking stimuli had a significant effect on amplitude discrimination ($p<0.001$). The average amplitude discrimination threshold was significantly different in all test conditions ($C1-C4$) and differed in the same order as in the frequency discrimination experiment (4.08 dB for $C4$, 3.60 dB for $C2$, 3.35 dB for $C3$, and 2.33 dB for $C1$). As expected, the amplitude thresholds for the no-masker condition ($C1$) never exceeded those for the one-masker condition ($C2$ or $C3$), which in turn were smaller than those for the two-masker condition ($C4$).

Changes in the amplitude discrimination thresholds due to masking are shown in Fig. 8. The average $(\Delta A)_0$ for $C1$ is subtracted from those for $C2-C4$ at the same reference frequency and the differences are plotted as bar graphs. The statistical significance ($p<0.05$) of the difference in $(\Delta A)_0$ is indicated by an asterisk. The masking effects were slightly more pronounced at 35 dB SL than at 20 dB SL ($p<0.01$). For both the low-frequency targets (F_{ref}=2 and 4 Hz) and the mid-frequency targets (F_{ref}=15 and 30 Hz), the effect of high-frequency maskers ($C3-C1$) was generally smaller than that in the other one-masker conditions ($C2-C1$). For
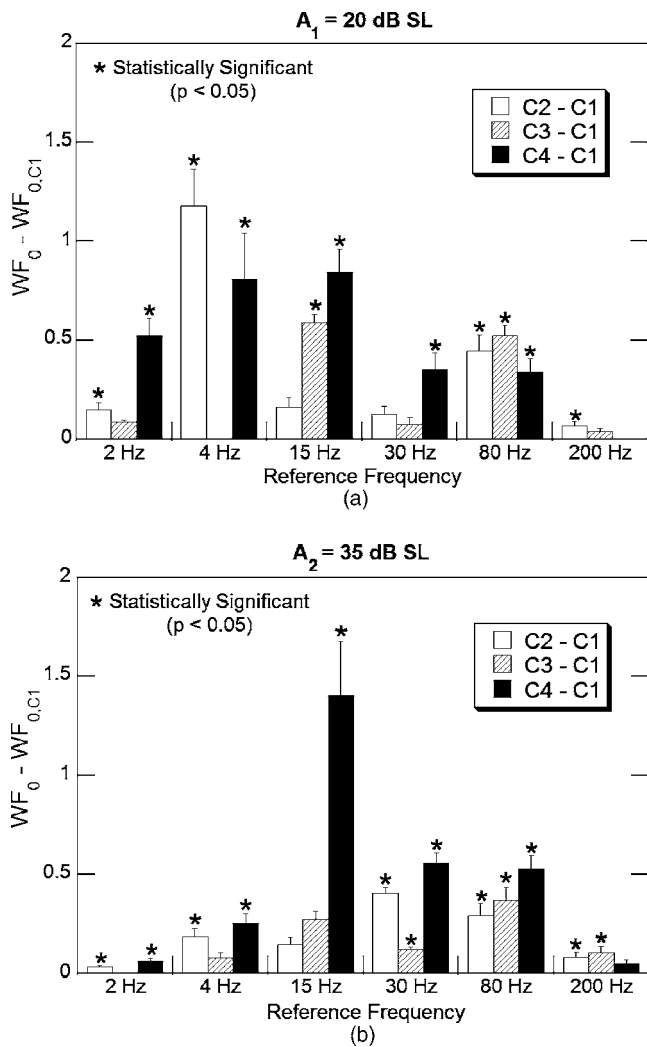
FIG. 5. Increase in frequency Weber fractions under the one- and two-masker conditions (C2–C4) for stimuli presented at (a) 20 dB SL and (b) 35 dB SL. WF_0 is the Weber fraction for the one- and two-masker conditions (C2–C4) and $WF_{0,C1}$ is the Weber fraction for the no-masker condition (C1). Differences are shown in a bar graph for each reference frequency. An asterisk indicates statistical significance ($p < 0.05$). Also shown are the standard errors.

the high-frequency targets ($F_{ref} = 80$ and 200 Hz), all the masking conditions affected the amplitude thresholds except for the 200-Hz targets at 20 dB SL for which the one-masker conditions ($C2 - C1$ and $C3 - C1$) did not significantly increase the amplitude threshold.

IV. DISCUSSION

The present study measured frequency and amplitude discrimination thresholds in the absence and presence of masking stimuli over a wide range of frequencies. In this section, we first compare our results from the no-masker condition to those in the literature, and then summarize the general effects of masking on frequency and amplitude discrimination thresholds. We also discuss the implications of our results in the context of tactile speech communication.

The frequency WFs in the no-masker condition (0.13–0.38 and 0.14–0.28 at 20 and 35 dB SL, respectively) are comparable to those reported by Goff (1967) (0.20–0.55 and

0.18–0.38 at 20 and 35 dB SL, respectively) and Rothenberg et al. (1977) (0.15–0.30 at 14 dB SL). The current data, however, do not show an increasing trend in WF with frequency as was apparent in the data of Goff (1967) and Rothenberg et al. (1977), although Rothenberg et al. (1977) did comment that some of their participants achieved very low WFs at high frequencies (0.05 and 0.03 at 150 and 200 Hz, respectively). As far as signal amplitude is concerned, the frequency WFs from the present study generally confirm the expected trend that WFs decrease with an increase in signal amplitude, presumably due to an increased signal-to-noise ratio in afferent signals at higher signal amplitudes (see Goff, 1967; LaMotte and Mountcastle, 1975). The one exception from the present study was that the frequency WFs for 200-Hz targets were significantly lower at 20 dB SL than at 35 dB SL. According to the psychophysical model proposed by Bolanowski Jr., et al. (1988); Gescheider et al., (2002); and Gescheider, Bolanowski, and Verrillo (2004), only the P channels respond to 200-Hz targets at 20 dB SL, and both the P and NP II channels are sensitive to 200-Hz targets at 35 dB SL. Our results at 200 Hz are consistent with previous studies that reported better tactual performance with high-frequency targets at intensities that activated only the P channels (Bensmaia et al., 2005; Horch, 1991; Summers et al., 1997). It is therefore possible that the activation of the NP II channels at 35 dB SL can somehow elevate frequency discrimination thresholds, although an explanation for such a phenomenon has yet to be discovered (although see Bensmaia et al., 2005; Bensmaia and Hollins, 2000; Summers et al., 1997; Verrillo and Gescheider, 1992).

The amplitude-discrimination thresholds obtained from the no-masker condition of the present study are similar to those reported in previous studies and vary with reference frequency and amplitude. The trend exhibited by the data depended on whether the target frequency was in the low-to-medium (2, 4, and 15 Hz) or medium-to-high (30, 80, and 200 Hz) regions. The amplitude discrimination thresholds in the lower frequency region were significantly higher at 20 dB SL than at 35 dB SL. This is consistent with the trend reported by Rinker et al. (1998) in that amplitude discrimination thresholds of 1-Hz waveforms decreased from roughly 1.3 to 0.8 dB as the waveform amplitude increased from 6.3 to 19.2 mm (a change of approximately 10 dB). The amplitude discrimination thresholds in the higher-frequency region did not change with amplitude in the present study. Several earlier studies have also reported that amplitude discrimination thresholds remained constant at amplitude levels above about 15 dB SL for mid- and high-frequency reference signals (Craig, 1972; Fucci et al., 1982; Gescheider et al., 1990; Knudsen, 1928; LaMotte and Mountcastle, 1975).

Although overall trends in the data were similar across participants, individual differences in threshold values were noted, particular for the masked conditions ($C2 - C4$). For frequency discrimination, no one subject stands out as consistently exhibiting the highest values of WF [e.g., see panel (d) of Fig. 4]; for amplitude discrimination, on the other hand, the performance of S4 is consistently worse than that of the other three participants [e.g., see panels (b) and (d) of

$A_1 = 20$ dB SL
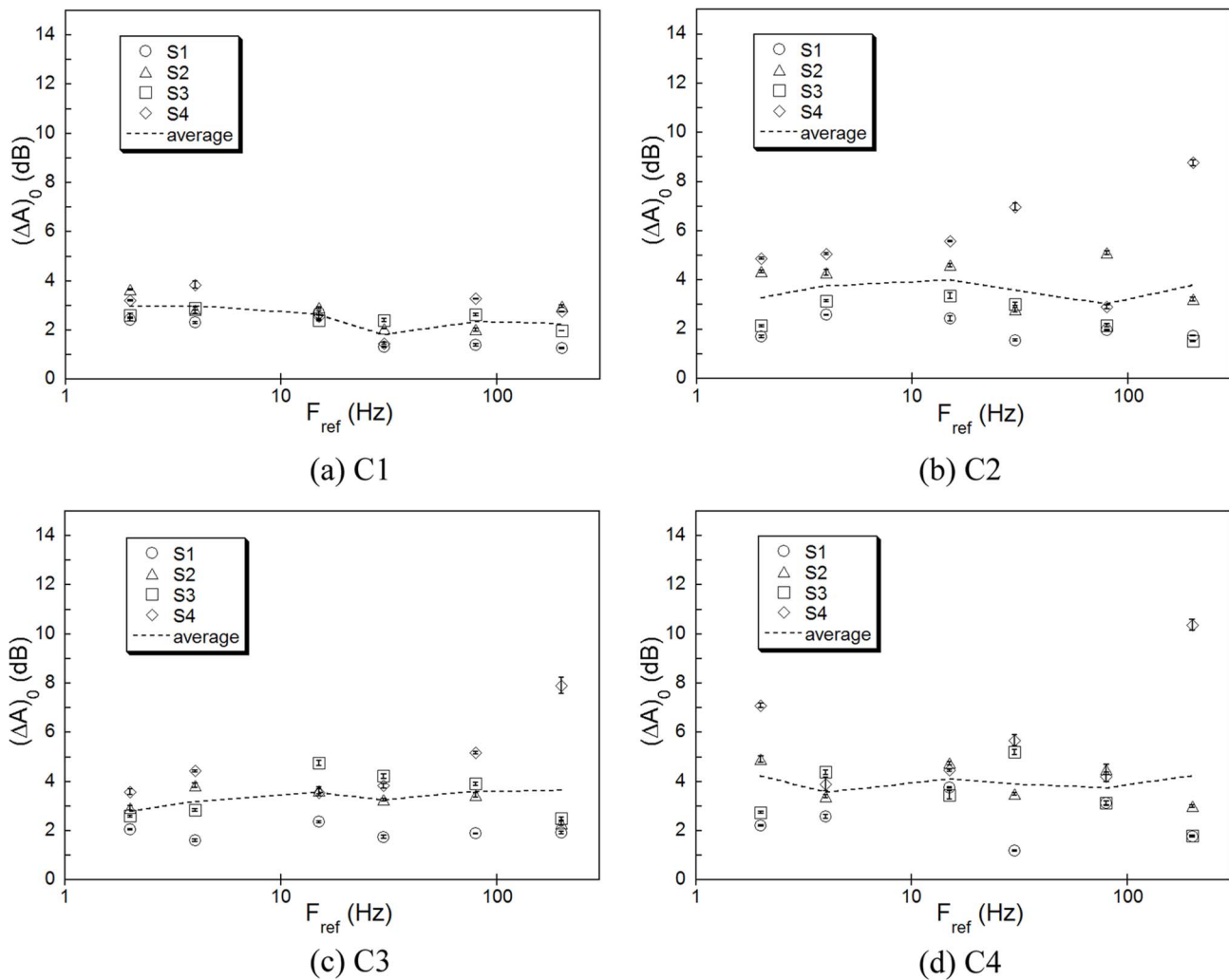
FIG. 6. Amplitude discrimination thresholds $(\Delta A)_0$ in decibels for the four participants as a function of reference frequency F_{ref} at 20 dB SL. Panel (a) shows the thresholds in the no-masker condition C1. Panels (b), (c), and (d) show thresholds for the one- and two-masker conditions C2, C3, and C4, respectively. Shown are individual participant's data (open symbols), the average (dashed lines), and standard errors.

Fig. 7]. One issue that could have potentially contributed to the significant interparticipant differences observed in the present study was the use of nominal detection threshold values for estimating the sensation levels of tactual signals. As mentioned earlier, the detection thresholds used in the present study were representative of three participants in Israr *et al.* (2004). Post detection thresholds of three participants in the present study showed that measured detection thresholds were within 6 dB from the nominal detection thresholds except for high-frequency targets where the thresholds varied up to 13 dB from the nominal detection thresholds (see the Appendix). From existing data in the literature, it appears that a small variation in the sensation level of tactual signals should not have a strong effect on discrimination thresholds. For example, Horch (1991) showed that the capability of humans to discriminate vibrations between 168 and 226 Hz did not change when the amplitude of the stimulus was varied from −4.6 dB to 3 dB relative to the reference. LaMotte and Mountcastle (1975) also demonstrated that frequency discrimination thresholds obtained at 30 Hz with roving amplitudes (24–36 dB SL) were identical

to those evaluated with equal subjective intensities, for both humans and monkeys. Therefore, the significant interparticipant differences were likely due to other factors such as familiarity with the experimental apparatus and training, and is not uncommon in psychophysical studies involving threshold testing (see, for example, Brisben, Hsiao, and Johnson, 1999). Previous studies have shown that training can lead to improved tactile discrimination performance, especially with new participants (Hnath-Chisolm and Medwetsky, 1988; Horch, 1991).

The detrimental effects of masking stimuli on frequency and amplitude discrimination found in the present study are consistent with similar findings from numerous earlier studies (Craig, 1972, 1974; Gescheider *et al.*, 1990; Gescheider, Verrillo, and Van Doren, 1982; Verrillo and Gescheider, 1992). In general, compared to the thresholds obtained in the no-masker condition, discrimination thresholds increased more under the two-masker condition (C4) than under the one-masker conditions (C2 and C3), and masking effects decreased as the frequency separation between the target and the maskers increased. The latter is consistent with findings
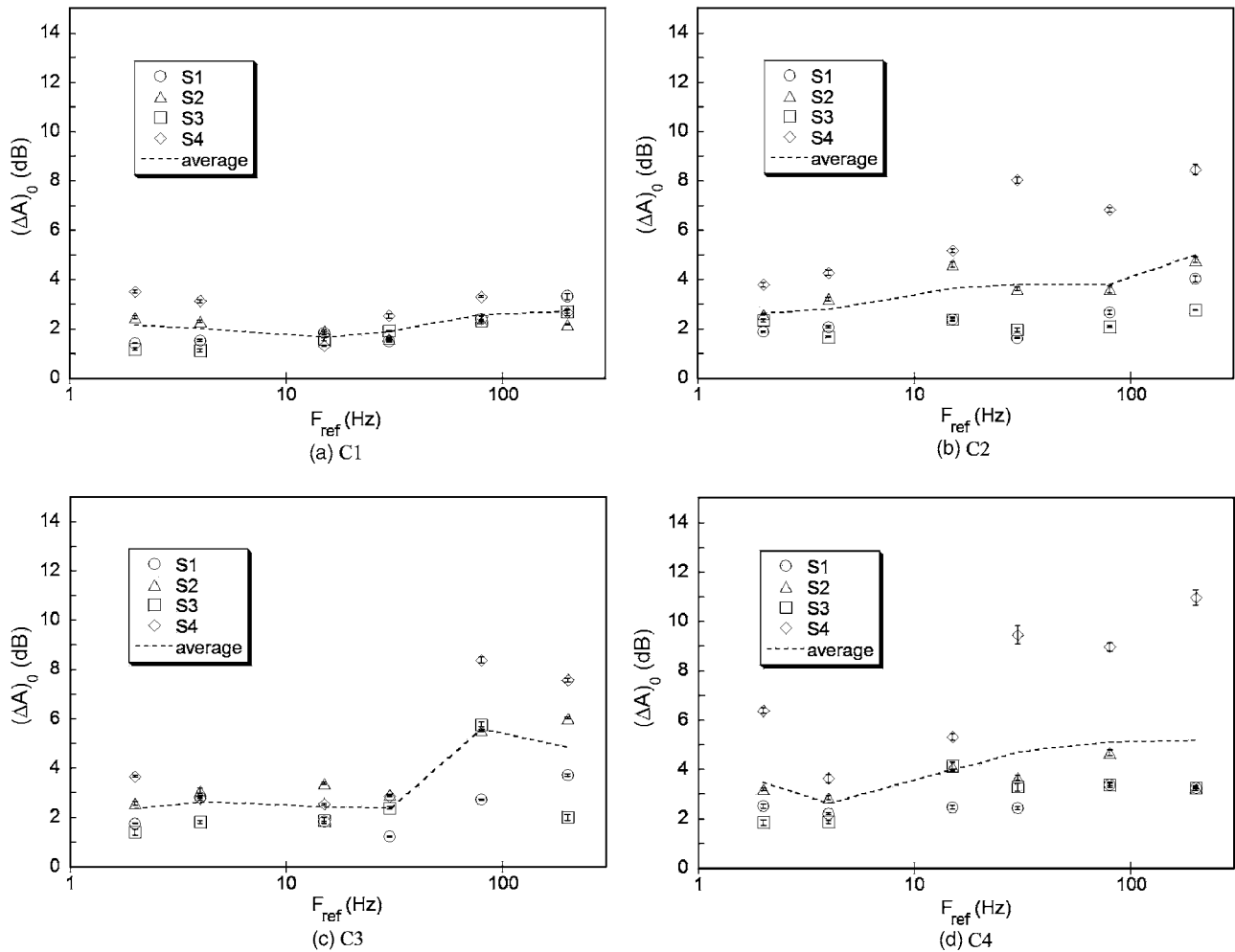
$A_2 = 35$ dB SL

FIG. 7. Amplitude discrimination thresholds $(\Delta A)_0$ in decibels for the four participants as a function of reference frequency F_{ref} at 35 dB SL. Panel (a) shows the thresholds in the no-masker condition C1. Panels (b), (c), and (d) show thresholds for the one- and two- masker conditions C2, C3, and C4, respectively. Shown are individual participant's data (open symbols), the average (dashed lines), and standard errors.

from other studies (Gescheider *et al.*, 2004; Hollins *et al.*, 1990; Summers *et al.*, 2005; Tan *et al.*, 1999, 2003). From the point of view of developing coding schemes for transmitting speech through a tactual display, one of the most important findings of the present study is that the discrimination of low-frequency targets is hardly affected by high-frequency maskers, and vice versa.

No definite model of frequency and amplitude discrimination is available, presumably due to the complex mechanisms underlying mechanical touch. Bolanowski Jr., *et al.* (1988) proposed a four-channel psychophysical model for the mechanical nature of touch in the dynamic range of about dc−400 Hz and 0−55 dB SL. According to this model for the glabrous skin of the thenar eminence, the detection thresholds of mechanoreceptors are contributed by four neural channels (NP I, NP II, NP III, and P) that partially overlap in their absolute sensitivity. In addition, each channel exhibits its own temporal and spatial characteristics (Verrillo and Gescheider, 1992). A similar psychophysical model was obtained at the fingertip by the same research group (Gescheider *et al.*, 2002, 2004). It is likely that suprathreshold stimuli activate more than a single channel (if not all four channels) as frequency and amplitude vary. The reference

and masker stimuli of the present study covers approximate the entire dynamic range of useful mechanical touch, in that the low-frequency targets are sensed by the NP I and NP III channels and also activate receptors for kinesthetic sensations at larger amplitudes, and the 200-Hz targets are sensed by the P channels as well as the NP II channels at higher amplitudes. Our results contribute to future studies of a model for frequency and amplitude discrimination in the presence of masking stimuli.

The present study was a continuation of our effort to use the tactual sense as a channel for speech communication. Despite decades of intense research efforts, performance with artificial hearing aids is still shadowed by that of natural methods such as the Tadoma method. Inspired by the Tadoma method, the TACTUATOR was developed to deliver both high-frequency low-amplitude vibrations and low-frequency high-amplitude movements to enrich the amount of speech information that can be transmitted through the tactual sense. Building on the author's previous studies using the TACTUATOR, the present study focused on user's ability to detect small variations in frequency or amplitude in the presence of a broadband speech signal. In this context, the main conclusions to be drawn from the present study are that low-

A_1 = 20 dB SL

★ Statistically Significant
(p < 0.05)

□ C2 – C1
▨ C3 – C1
■ C4 – C1

(a)

A_2 = 35 dB SL

★ Statistically Significant
(p < 0.05)

□ C2 – C1
▨ C3 – C1
■ C4 – C1

(b)
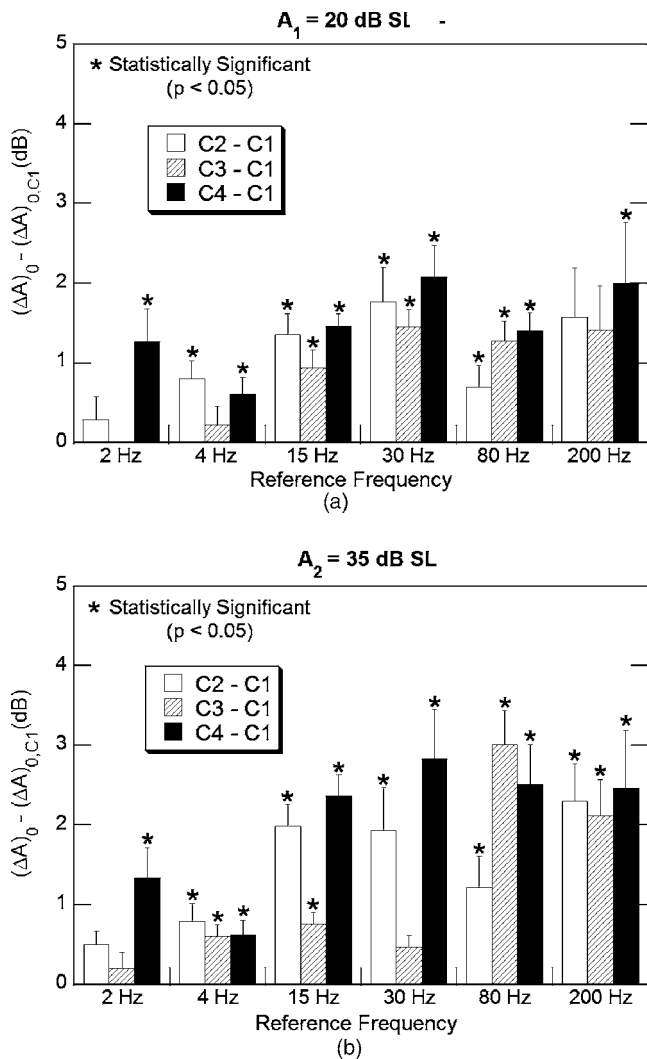
FIG. 8. Increase in amplitude discrimination thresholds in (decibels) under the one- and two-masker conditions (C2–C4) for stimuli presented at (a) 20 dB SL and (b) 35 dB SL. $(\Delta A)_0$ is the amplitude discrimination threshold for the one- and two-masker conditions (C2–C4) and $(\Delta A)_{0,C1}$ is the amplitude discrimination threshold for the no-masker condition (C1). Differences are shown in a bar graph for each reference frequency. An asterisk indicates statistical significance ($p < 0.05$). Also shown are the standard errors.

and high-frequency tactual signals do not interfere with each other, that experience with the apparatus or task improved discrimination performance, and that tactual discrimination thresholds appear to be too large for encoding more than a few speech features. The relative independence of the low- and high-frequency tactual channels will be exploited in the near future to improve the transmission of one speech cue: place of articulation. Previous studies on tactile aids were able to provide manner and voicing distinctions, but failed to transmit the place of articulation cue (Bernstein *et al.*, 1991; Carney *et al.*, 1993; Clements, Braida, and Durlach, 1982, 1988; Weisenberger, Craig, and Abbott, 1991; Weisenberger and Percy, 1995). Place of articulation is well correlated with the frequency values of the first two formats *F*1 and *F*2 and their transitions (see, for example, Ainsworth, 1968; Ali, Van der Spiegel, and Mueller, 2001a, 2001b; Jongman and Miller, 1991; Jongman, Wayland, and Wong, 2000; Sharf and Hemeyer, 1972; Stevens and Klatt, 1974). Based on the frequency and amplitude discrimination thresholds found in the present

TABLE II. Nominal and individual detection thresholds for S1, S2, and S3.

Frequency (Hz)	Detection thresholds (dB re 1 μm peak)			
	Nominal	S1	S2	S3
2	38	39	41	33
4	34	34	35	32
15	25	29	26	24
30	16	19	22	21
80	−8	−16	5	4
200	−18	−22	−9	−12

study, the human hand does not have the resolution required for discriminating changes in high-frequency vibrations in the *F*1 and *F*2 region. A direct presentation of vibration signals associated with *F*1 and *F*2 is therefore expected to lead to poor speech reception performance. We will use a low-frequency motional signal to supplement the *F*1 and *F*2 vibrations. The direction (extension versus flexion of a finger) and the extent of the motional signals will be used to indicate to a user whether the formats increase or decrease and how fast they are changing. We hope to demonstrate that the strategy of redundantly encoding place of articulation cues via the simultaneous use of low- and high-frequency signals will eventually contribute toward improved speech transmission through the sense of touch.

ACKNOWLEDGMENTS

APPENDIX

To check the validity of the threshold levels shown in Table I, detection thresholds for *S*1, *S*2, and *S*3 were measured after the main experiments and are presented in Table II. S4 was no longer available for the additional testing.

[1]There exists another device that can display both high-frequency vibrations and low-frequency motions with appreciable amplitude (Eberhardt *et al.*, 1994). However, it either operates closed loop under 20 Hz or open loop above 100 Hz, but does not deliver signals in the frequency range 20–100 Hz.
[2]The first TACTUATOR was developed at MIT (Tan and Rabinowitz, 1996). A second device, the TACTUATORII, was then developed at Purdue University with essentially the same hardware, but a new two degrees-of-freedom controller (Israr *et al.*, 2004).
[3]The actual noise level of the sensor (due to mechanical and electrical noises) can be found in Fig. 8 of Tan and Rabinowitz (1996).
[4]To estimate the displacements associated with the sensation levels discussed in this paragraph, please see Table I for the detection thresholds at six frequencies tested in the present study.
[5]To examine training effects, participants *S*1 and *S*2 completed two runs of experimental conditions. For both subjects, frequency-discrimination thresholds improved in the second run. In the amplitude discrimination experiments, thresholds for *S*1 were not significantly different across the two runs, while those of *S*2 were better in the first run. The better of the two runs was used in the final data analyses. Because of the mixed training trends observed here, *S*3 and *S*4 completed only one set of runs.

Ainsworth, W. A. (**1968**). "First formant transitions and the perception of synthetic semivowels," J. Acoust. Soc. Am. **44**(3), 689–694.

Ali, A. M. A., Van der Spiegel, J., and Mueller, P. (**2001a**). "Acoustic-phonetic features for the automatic classification of fricatives," J. Acoust. Soc. Am. **109**(5), 2217–2235.

Ali, A. M. A., Van der Spiegel, J., and Mueller, P. (**2001b**). "Acoustic-phonetic features for the automatic classification of stop consonants," IEEE Trans. Speech Audio Process. **9**(8), 833–841.

Bensmaia, S. J., and Hollins, M. (**2000**). "Complex tactile waveform discrimination," J. Acoust. Soc. Am. **108**(3), 1236–1245.

Bensmaia, S., Hollins, M., and Yau, J. (**2005**). "Vibrotactile intensity and frequency information in the pacinian system: A psychophysical model," Percept. Psychophys. **67**(5), 828–841.

Bernstein, L. E., Demorest, M. E., Coulter, D. C., and O'Connell, M. P. (**1991**). "Lipreading sentences with vibrotactile vocoders: Performance of normal-hearing and hearing-impaired Subjects," J. Acoust. Soc. Am. **90**(6), 2971–2984.

Bolanowski Jr., S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (**1988**). "Four channels mediate the mechanical aspects of touch," J. Acoust. Soc. Am. **84**(5), 1680–1694.

Brisben, A. J., Hsiao, S. S., and Johnson, K. O. (**1999**). "Detection of vibration transmitted through an object grasped in the hand," J. Neurophysiol. **81**, 1548–1558.

Carney, A. E., Osberger, M. J., Carney, E., Robbins, A. M., Renshaw, J., and Miyamoto, R. T. (**1993**). "A comparison of speech discrimination with cochlear implants and tactile aids," J. Acoust. Soc. Am. **94**(4), 2036–2049.

Clements, M. A., Braida, L. D., and Durlach, N. I. (**1982**). "Tactile communication of speech: II. Comparison of two spectral displays in a vowel discrimination task," J. Acoust. Soc. Am. **72**(4), 1131–1135.

Clements, M. A., Braida, L. D., and Durlach, N. I. (**1988**). "Tactile communication of speech: Comparison of two computer-based displays," J. Rehabil. Res. Dev. **25**(4), 25–44.

Craig, J. C. (**1972**). "Difference threshold for intensity of tactile stimuli," Percept. Psychophys. **11**, 150–152.

Craig, J. C. (**1974**). "Vibrotactile difference thresholds for intensity and the effect of a masking stimulus," Percept. Psychophys. **15**, 123–127.

Eberhardt, S. P., Bernstein, L. E., Barac-Cikoja, D., Coulter, D. C., and Jordan, J. (**1994**). "Inducing dynamic haptic perception by the hand: System description and some results," *Proceedings of the American Society of Mechanical Engineers: Dynamic Systems and Control*, Vol. **55**, (The American Society of Mechanical Engineers, NY), pp. 345–351.

Franzen, O., and Nordmark, J. (**1975**). "Vibrotactile frequency discrimination," Percept. Psychophys. **17**(5), 480–484.

Fucci, D., Small, L. H., and Petrosino, L. (**1982**). "Intensity difference limens for lingual vibrotactile stimuli," J. Magn. Reson. **1**, 54–56.

Geldard, F. A. (**1957**). "Adventures in tactile literacy," Am. Psychol. **12**, 115–124.

Gescheider, G. A., Bolanowski, S. J., and Verrillo, R. T. (**2004**). "Some characteristics of tactile channels," Behav. Brain Res. **148**(1), 35–40.

Gescheider, G. A., Bolanowski, S. J., Pope, J. V., and Verrillo, R. T. (**2002**). "A four-channel analysis of the tactile sensitivity of the fingertip: Frequency selectivity, spatial summation, and temporal summation," Somatosens Mot Res. **19**(2), 114–124.

Gescheider, G. A., Bolanowski Jr., S. J., Verrillo, R. T., Arpajian, D. J., and Ryan, T. F. (**1990**). "Vibrotactile intensity discrimination measured by three methods," J. Acoust. Soc. Am. **87**(1), 330–338.

Gescheider, G. A., Verrillo, R. T., and Van Doren, C. L. (**1982**). "Prediction of vibrotactile masking functions," J. Acoust. Soc. Am. **72**(5), 1421–l426.

Gescheider, G. A., Zwislocki, J. J., and Rasmussen, A. (**1996**). "Effects of stimulus duration on the amplitude difference limen for vibrotaction," J. Acoust. Soc. Am. **100**(4), 2312–2319.

Globe, A. K., and Hollins, M. (**1993**). "Vibrotactile adaptation enhances amplitude discrimination," J. Acoust. Soc. Am. **93**(1), 418–424.

Globe, A. K., and Hollins, M. (**1994**). "Vibrotactile adaptation enhances frequency discrimination," J. Acoust. Soc. Am. **96**(2), 771–780.

Goff, G. D. (**1967**). "Differential discrimination of frequency of cutaneous mechanical vibration," J. Exp. Psychol. **74**(2), 294–299.

Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d' for M-alternative forced choice," Percept. Psychophys. **26**(2), 168–170.

Hnath-Chisolm, T., and Medwetsky, L. (**1988**). "Perception of frequency contours via temporal and spatial tactile transforms," Ear Hear. **9**(6), 322–328.

Hollins, M., Globe, A. K., Whitsel, B. L., and Tommerdahl, M. (**1990**). "Time course and action spectrum of vibrotactile adaptation," Somatosens Mot Res. **7**(2), 205–221.

Horch, K. (**1991**). "Coding of vibrotactile stimulus frequency by Pacinian corpuscle afferents," J. Acoust. Soc. Am. **89**(6), 2827–2836.

Israr, A., Meckl, P. H., and Tan, H. Z. (**2004**). "A two DOF controller for a multi-finger tactual display using a loop-shaping technique," *Proceedings of the 2004 ASME International Mechanical Engineering Congress*, Anaheim, CA, Nov. 13–19 (The American Society of Mechanical Engineers, NY), pp. 1083–1089.

Jongman, A., and Miller, J. D. (**1991**). "Method for the location of burst-onset spectra in the auditory-perceptual space: A study of place of articulation in voiceless stop consonants," J. Acoust. Soc. Am. **89**(2), 867–873.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of english fricatives," J. Acoust. Soc. Am. **108**(3), 1252–1263.

Knudsen, V. O. (**1928**). ""Hearing" with the sense of touch," J. Gen. Psychol. **1**, 320–352.

LaMotte, R. H., and Mountcastle, V. B. (**1975**). "Capacities of humans and monkeys to discriminate vibratory stimuli of different frequency and amplitude: A correlation between neural events and psychological measurements," J. Neurophysiol. **38**(3), 539–559.

Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**(8), 1279–1292.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467–477.

Reed, C. M., and Durlach, N. I. (**1998**). "Note on information transfer rates in human communication," Presence—Teleoperators & Virtual Environments **7**(5), 509–518.

Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Braida, L. D., Conway-Fithian, S., and Schultz, M. C. (**1985**). "Research on the tadoma method of speech communication," J. Acoust. Soc. Am. **77**(1), 247–257.

Rinker, M. A., Craig, J. C., and Bernstein, L. E. (**1998**). "Amplitude and period discrimination of haptic stimuli," J. Acoust. Soc. Am. **104**(1), 453–463.

Rothenberg, M., Verrillo, R. T., Zahorian, S. A., Brachman, M. L., and Bolanowski Jr., S. J. (**1977**). "Vibrotactile frequency for encoding a speech parameter," J. Acoust. Soc. Am. **62**, 1003–1012.

Sharf, D. J., and Hemeyer, T. (**1972**). "Identification of place of consonant articulation from vowel formant transitions," J. Acoust. Soc. Am. **51**(2), 652–658.

Stevens, K. N., and Klatt, D. H. (**1974**). "Role of formant transitions in the voiced-voiceless distinction for stops," J. Acoust. Soc. Am. **55**(3), 653–659.

Summers, I. R., Cooper, P. G., Wright, P., Gratton, D. A., Milnes, P., and Brown, B. H. (**1997**). "Information from time-varying vibrotactile stimuli," J. Acoust. Soc. Am. **102**(6), 3686–3696.

Summers, I. R., Whybrow, J. J., Gratton, D. A., Milnes, P., Brown, B. H., and Stevens, J. C. (**2005**). "Tactile information transfer: A comparison of two stimulation sites," J. Acoust. Soc. Am. **118**(4), 2527–2534.

Tan, H. Z., Durlach, N. I., Reed, C. M., and Rabinowitz, W. M. (**1999**). "Information transmission with a multifinger tactual display," Percept. Psychophys. **61**(6), 993–1008.

Tan, H. Z., and Rabinowitz, W. M. (**1996**). "*A new multi-finger tactual display*," *Proceedings of the International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, edited by K. Danai, Vol. **58** (The American Society of Mechanical Engineers, NY), pp. 515–522.

Tan, H. Z., Reed, C. M., Delhorne, L. A., Durlach, N. I., and Wan, N. (**2003**). "Temporal masking of multidimensional tactual stimuli," J. Acoust. Soc. Am. **114**(6), 3295–3308.

Verrillo, R. T., and Gescheider, G. A. (**1992**). "*Perception via the sense of touch*," *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr Publishers, London), pp. 1–36.

Weisenberger, J. M., Craig, J. C., and Abbott, G. D. (**1991**). "Evaluation of a principal-components tactile aid for the hearing-impaired," J. Acoust. Soc. Am. **90**(4), 1944–1957.

Weisenberger, J. M., and Percy, M. E. (**1995**). "The transmission of phoneme-level information by multichannel tactile speech perception aids," Ear Hear. **16**(4), 392–406.

Yuan, H., Reed, C. M., and Durlach, N. I. (**2005a**). "Temporal onset-order discrimination through the tactual sense," J. Acoust. Soc. Am. **117**(5), 3139–3148.

Yuan, H., Reed, C. M., and Durlach, N. I. (**2005b**). "Tactual display of consonant voicing as a supplement to lipreading," J. Acoust. Soc. Am. **118**(2), 1003–1015.

The mean matters: Effects of statistically defined nonspeech spectral distributions on speech categorization

Lori L. Holt[a)]

Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Adjacent speech, and even nonspeech, contexts influence phonetic categorization. Four experiments investigated how preceding sequences of sine-wave tones influence phonetic categorization. This experimental paradigm provides a means of investigating the statistical regularities of acoustic events that influence online speech categorization and, reciprocally, reveals regularities of the sound environment tracked by auditory processing. The tones comprising the sequences were drawn from distributions sampling different acoustic frequencies. Results indicate that whereas the mean of the distributions predicts contrastive shifts in speech categorization, variability of the distributions has little effect. Moreover, speech categorization is influenced by the global mean of the tone sequence, without significant influence of local statistical regularities within the tone sequence. Further arguing that the effect is strongly related to the average spectrum of the sequence, notched noise spectral complements of the tone sequences produce a complementary effect on speech categorization. Lastly, these effects are modulated by the number of tones in the acoustic history and the overall duration of the sequence, but not by the density with which the distribution defining the sequence is sampled. Results are discussed in light of stimulus-specific adaptation to statistical regularity in the acoustic input and a speculative link to talker normalization is postulated. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354071]

I. INTRODUCTION

The fluency of speech communication makes it a rarity for any particular speech sound to be present without context. Many studies have confirmed that the nature of adjacent context greatly influences how speech is categorized (for a review, Repp, 1982). For example, following /al/, listeners identify speech targets more often as "ga" than "da" whereas, following /ar/, the same speech targets are more often identified as "da" (Mann, 1980). Lotto *et al.* (1997) found that Japanese quail (*Coturnix coturnix japonica*) trained to peck a key in response to /ga/ or /da/ exhibit context-dependent responses when /al/ or /ar/ precede novel test tokens along a /ga/ to /da/ series. Quail trained to peck to /ga/ respond more vigorously when /al/ precedes novel test syllables. Thus, it appears that context-dependent response to speech is not species specific, implicating instead rather general auditory processing.

Results with human adult listeners support this possibility. When nonspeech sine wave tone contexts with spectra modeling the third-formant (F3) offset frequency of /al/ and /ar/ precede /ga/ to /da/ targets, context-dependent speech categorization shifts are observed that mirror the effects produced by /al/ and /ar/ speech contexts. Other studies have verified the generality of this finding (for a brief review, see Lotto and Holt, 2006). Nonspeech context effects on speech categorization have been observed with nonspeech analogs of consonant contexts affecting categorization of consonants (Fowler *et al.*, 2000; Lotto *et al.*, 2003; Lotto and Kluender,

1998) and vowels (Holt *et al.*, 2000). Nonspeech models of vowel contexts also influence consonant targets (Coady *et al.*, 2003; Holt, 1999). Steady-state and frequency-modulated sine-wave tones, single formant harmonic stimuli, and band-pass noise have each been effective as nonspeech contexts in shifting speech categorization. Moreover, the interactions of speech and nonspeech signals are reciprocal; speech precursors also influence perception of nonspeech targets (Stephens and Holt, 2003). Finally, like observed context dependencies among speech contexts and targets, later-occurring nonspeech contexts influence perception of preceding speech (Wade and Holt, 2005a). Language-independent auditory processes may also contribute to compensation for phonological assimilation (Gow, 2003; Mitterer *et al.*, 2006).

Thus, there is mounting evidence that speech is perceived relative to the characteristics of neighboring sounds, whether speech or nonspeech, and therefore without respect to whether the signals originate from a vocal-tract source (see also Holt, 1999; Lotto and Kluender, 1998). When nonspeech contexts possess spectra similar to the speech contexts they model, they are effective in producing effects of context on speech categorization. This perceptual pattern has been described as spectral contrast in that higher-frequency context stimuli (e.g., /al/ or the nonspeech tones that model it) shift perception toward lower frequency targets (i.e., /ga/; Lotto and Kluender, 1998). It does not appear that speech-specific information is necessary to elicit context-dependent phonetic categorization. Rather, the body of evidence suggests a role for more general perceptual mechanisms. These

---

a)Electronic mail: lholt@andrew.cmu.edu

studies thus serve as an example of how studies of context effects on phonetic categorization can be informative about speech processing.

Reciprocally, the generality of phonetic context effects suggests it may be possible to use context-dependent phonetic categorization as a means of understanding auditory processing, generally speaking. Using listeners' context-dependent phonetic categorization as an index, it is possible to investigate how different signals interact in auditory processing. In this way, phonetic categorization can serve as a platform from which to address general perceptual questions via manipulation and control of the context stimuli. The present experiments exploit this possibility.

A. Auditory sensitivity to acoustic distributions

It is now quite well established that auditory processing is sensitive to a variety of regularities within the acoustic environment. Human adult (Saffran *et al.*, 1996), infant (Saffran *et al.*, 1996), and nonhuman (Hauser *et al.*, 2001) listeners are able to use transitional probabilities to segment running speech streams as well as streams of nonspeech acoustic elements (Saffran *et al.*, 1999). Adult and infant listeners appear to be sensitive to frequency distributions of phonetic exemplars (Maye *et al.*, 2002). Video-game players are able to extract higher-dimensional acoustic cues in aid of auditory categorization in the course of game play (Wade and Holt, 2005b). In each of these cases, experienced regularity in the acoustic input shapes subsequent auditory perception, including phonetic perception. However, within the environment, there are an infinite number of statistics that might be tracked by perceptual systems. Therefore, it is of much interest to understand the perceptual and cognitive constraints that may be imposed upon the processing of regularity from the perceptual world and the learning that may result from it. Because novel nonspeech acoustic contexts can be so readily defined in entirely statistical terms, their influence on phonetic categorization affords an opportunity to carefully manipulate and control the regularity in the acoustic environment and to observe the effects of this regularity on speech processing. The present experiments make use of the effect of nonspeech contexts on phonetic categorization as a means by which to investigate the bounds of perceptual sensitivity to statistical regularity in the input in an effort to zero in on the regularities in acoustic context that are important for online speech perception.

For example, is auditory processing sensitive only to interactions among temporally adjacent acoustic events as in the examples above (Lotto and Kluender, 1998; Holt *et al.*, 2000)? Or, rather, is it also responsive to distributional regularities among multiple acoustic events that unfold across time? Taking a hint from speech processing, there is some long-standing evidence that may be characteristic of such sensitivity. In a classic study, Ladefoged and Broadbent (1957) manipulated the formant frequencies of a sentence preceding a target word that could be perceived as *bit*, *bet*, *bat*, or *but*. These manipulations to the precursor sentence affected the way listeners categorized the final target word, producing context effects such that the target word was iden-

tified differently as the formant frequencies of the preceding sentence changed. The spectral manipulations to the formant frequencies of the precursor sentence created a perceived change in voice characteristics, making it sound like different speakers uttered the sentence. Accordingly, these results have been interpreted as an instance of perceptual normalization for speaker characteristics (e.g., Repp, 1982). However, the overall distributions of spectral energy in the precursor sentence also shifted as the formant frequencies were manipulated. Thus, the effect of precursor on identification of the final word might be described also as an effect of varying distributions of spectral energy on perception of the final word. If listeners are sensitive to the long-term regularity of the spectral distributions of acoustic energy, it may affect phonetic categorization.

There is some evidence that supports this latter possibility. Watkins and Makin (1994) conducted a series of studies similar to those of Ladefoged and Broadbent (1957), demonstrating that the effect of preceding sentence on speech categorization persists even when the context sentence is played backwards and even when it is replaced by noise that matches the long-term spectra of the context sentence. These results are consistent with findings on perceptual compensation for spectral distortion in transmission channels (Darwin *et al.*, 1989; Kiefte and Kluender, 2001; Watkins, 1991; Watkins and Makin, 1996) in which categorization of a speech target is shifted by filtering the context sentence. The target sound is heard as if it were filtered with the inverse of the filter shape applied to the context. They are also consistent with recent reports of the effects of context on perceptual normalization of lexical tone (Francis *et al.*, 2006). In all of these experiments, perception of the target sound appears to be relative to and contrastive with context. This is precisely the pattern of perception required for producing the effects observed by Ladefoged and Broadbent and presumed to relate to speaker normalization.

Extending the principle that context-dependent phonetic categorization may be a useful index of the types of acoustic information of significance in auditory processing and, taking a cue from the influential findings of Ladefoged and Broadbent (1957), Holt (2005) tested explicitly whether listeners are sensitive to statistically defined distributions of acoustic energy. Like the Ladefoged and Broadbent (1957) study, these experiments manipulated the spectral distribution of energy across sentence-length precursors. However, the spectral distributions were sampled not with shifted formant frequencies, but with purely nonspeech sine wave tones. As such, there was no information for any particular speaker or vocal source or any information about speech whatsoever. The question of interest was whether phonetic categorization of speech syllable targets shifts as a function of the characteristics of the spectral distributions sampled by the sequences of sine wave tones. If so, it would suggest that auditory processing is sensitive to statistical distributions of acoustic energy that unfold across time and that, moreover, this sensitivity can play a role in phonetic categorization—perhaps accounting for some of the perceptual shifts that have been attributed to speaker normalization processes.

In these experiments, 22 70-ms sine-wave tones were

FIG. 1. Stimulus structure and results of experiment 1. (a) Schematic illustration of the experiment 1 stimulus structure based on the paradigm of Holt (2005); (b) a representive stimulus spectrogram from each of the eight (2 Mean × 4 Variability) conditions; (c) listeners' average "ga" categorization responses to the /ga/ to /da/ series stimuli in the context of preceding high and low mean acoustic histories with 1 tone at the mean frequency, 3 tones sampling 100 Hz around the mean frequency, 7 tones sampling 300 Hz around the mean frequency, or 21 tones sampling 1000 Hz around the mean frequency; and (d) average "ga" categorization responses collapsed across the target stimulus series as a function of the distribution characteristics of the acoustic history precursors.

presented sequentially across 2.2 s. These tones were followed by a speech target stimulus drawn from a series ranging perceptually from /ga/ to /da/. Figure 1(a) depicts the construction of the stimuli. Twenty-one tones formed an "acoustic history" that preceded the speech targets. The frequencies of these tones were drawn from distributions that differentially sampled the acoustic spectrum such that spectral distribution characteristics differentiated conditions. A low mean condition with an 1800-Hz mean sampled 1300–2300 Hz. A high mean condition with a 2800-Hz mean sampled 2300–3300 Hz. These distribution means modeled the spectral characteristics of the tones Lotto and Kluender (1998) found to be effective in shifting listeners' /ga/-/da/ speech categorization responses. Within the acoustic history distributions, tones were equally probable (presented once per stimulus) and were sampled in 50-Hz steps. The tones of the acoustic histories were presented sequentially;

30 ms of silence separated tones. The sequence of tones comprising the acoustic history was reordered on each trial of the experiment; thus, each trial was a unique stimulus and idiosyncrasies of particular tone orderings could not account for any observed effects on phonetic categorization. Finally, each acoustic history ended in a 2300-Hz standard tone that was found, in a pilot experiment, to have no effect on labeling the /ga/ to /da/ targets (Holt, 2005). This prevented any observed context effects from originating from adjacent interactions of acoustic stimuli such as those described by Lotto and Kluender (1998).

Holt (2005, 2006) reports that these statistically defined acoustic contexts exert a strong contrastive influence on phonetic categorization. Sequences of tones drawn from the high mean acoustic history distribution result in more "ga" responses (i.e., the target characterized by greater low-frequency energy) whereas the same speech targets are more often labeled as "da" when preceded by tone sequences drawn from the low mean distribution. Further studies demonstrate that the influence of the acoustic histories does not depend on temporal adjacency with the speech target. Even when the acoustic histories are separated from the speech targets by 1.3 s of silence or by 13 repetitions of the standard tone (across 1.3 s), the acoustic histories elicit a context effect on speech categorization (Holt, 2005). For these effects to emerge, the auditory system must exhibit sensitivity to the distributional characteristics of acoustic events emerging across time. The precise characteristics of the sequences varied on a trial-by-trial basis; the distribution characterizes the influence of the acoustic histories on speech categorization. The implication of these studies is that the auditory system may "tune" or calibrate its response to spectral characteristics of the acoustic environment such that speech processing is thoroughly context dependent and relative to the distribution of acoustic energy present across some preceding temporal window, even if the energy across that window is nonspeech.[1]

The Holt (2005) results implicate the distribution mean as an important determiner of the influence of statistical regularity on auditory perceptual processing. However, it is not yet known the extent to which other distribution characteristics play a role. The present experiments extend the principle of using phonetic context effects as an empirical index of auditory processing to examine the kinds of distributional information to which listeners exhibit sensitivity in online speech processing. Experiment 1 investigates whether listeners are sensitive to distribution variance by examining whether phonetic context effects produced by the acoustic history distribution means are modulated also by the variability of the distributions across the spectrum. Experiment 2 seeks to understand what information contributes to the overall distribution, specifically whether the auditory system exhibits sensitivity to local as well as global regularities in acoustic history distributions. The results of each of the first two experiments underscore the significance of the mean of the distribution of spectral energy in predicting effects on phonetic categorization. Therefore, experiment 3 explicitly tests the consequent hypothesis that spectral complements of the acoustic histories should produce a complementary effect

Stimulus 1   **Stimulus 3**   **Stimulus 5**   **Stimulus 7**   **Stimulus 9**
/ga/   **/da/**
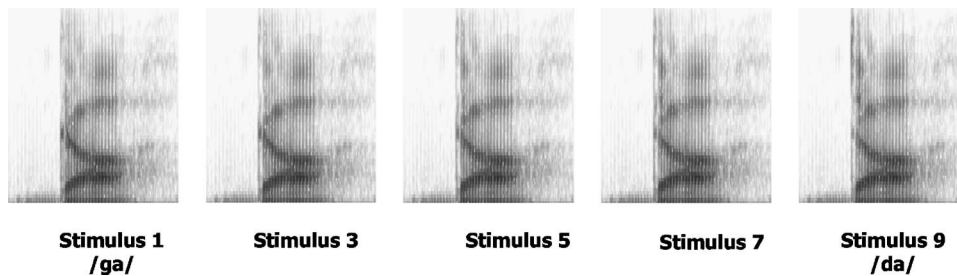
FIG. 2. Spectrograms of five representative speech target stimuli along the nine-step series.

on phonetic categorization. Finally, in two parts, experiment 4 attempts to tease apart how much information the auditory system requires to exhibit sensitivity to a distribution and whether the sampling density and/or duration of this information modulates its effect on phonetic categorization. Each of the experiments manipulates the distribution characteristics of the nonspeech acoustic history precursors on phonetic categorization to attempt to understand the boundaries of and constraints on online auditory processing of statistical regularity in the acoustic environment. The experiments address multiple questions of how distribution characteristics of context stimuli influence phonetic categorization. They are bound together in that their results converge on the conclusion that the mean of the distribution of preceding acoustic energy matters in auditory processing. Statistical regularity in the acoustic environment is tracked and employed in evaluating incoming acoustic signals; acoustic signals appear to be perceived relationally and contrastively with respect to the distribution mean of preceding acoustic energy. For each of the observed effects, listeners' speech categorization is robustly influenced by the mean of the acoustic history distribution of nonspeech sounds that precede the speech target.

II. EXPERIMENT 1

In investigating which types of acoustic regularity are important for speech processing, distribution variability is an obvious candidate. The results reported by Holt (2005) hint that distribution variability may be important in modulating the context effects on speech categorization produced by acoustic histories. In addition to the high (2800 Hz) and low (1800 Hz) mean experimental conditions, Holt (2005) tested two control conditions with middle mean acoustic history means of 2300 Hz preceding the speech targets. The two control conditions differed from one another only in the variability of the distributions from which the tones composing them were drawn. One condition sampled a frequency distribution equivalent to the 1000-Hz ranges sampled by the high and low mean experimental conditions. The other sampled the joint range of the high and low mean conditions taken together, 2000 Hz. Therefore, the variability of the distributions distinguished the two control conditions. Holt reported a small, statistically insignificant shift in speech categorization as a function of the distribution variability such that there were somewhat more "ga" responses when the middle mean acoustic histories had greater distribution variability. It is unclear whether this shift was perceptually meaningful as it was considerably smaller than the shifts observed as a

function of distribution mean. Therefore, experiment 1 explicitly manipulates the variability of the sound distributions over a greater frequency range and across the experimental high and low mean conditions to explicitly test whether acoustic history distribution variability influences speech categorization.

A. Methods

1. Participants

Ten volunteers from Carnegie Mellon University participated for course credit or a small payment. All listeners reported normal hearing and were monolingual native English speakers.

2. Stimuli

Stimulus design is illustrated in Fig. 1(a). For each stimulus an acoustic history composed of 22 sine-wave tones [21-tone acoustic history plus standard tone at 2300 Hz, as in the Holt (2005) experiments] preceded a 50-ms silent interval and a speech target drawn from a stimulus series varying perceptually from /ga/ to /da/.

a. Speech targets. The nine speech target stimuli were identical to those reported in earlier studies (Holt, 2005, 2006; Wade and Holt, 2005a). The stimuli were derived from natural /ga/ and /da/ recordings from a monolingual male native English speaker (Computer Speech Laboratory, Kay Elemetrics, Lincoln Park, NJ; 20-kHz sampling rate, 16-bit resolution). Multiple natural productions of the syllables were recorded and, from this set, one /ga/ and one /da/ token were selected that were nearly identical in spectral and temporal properties except for the onset frequencies of F2 and F3. LPC analysis was performed on each of these tokens and a nine-step series of filters was created to span these endpoints (Analysis-Synthesis Laboratory, Kay Elemetrics) such that the onset frequencies of F2 and F3 varied approximately linearly between /g/ and /d/ endpoints. These filters were excited by the LPC residual of the original /ga/ production to create an acoustic series spanning the natural /ga/ and /da/ endpoints in approximately equal steps. The series was judged by the experimenter to shift between natural-sounding /ga/ and /da/ tokens and this impression was confirmed by regular shifts in phonetic categorization across the series by participants in previous studies (Holt, 2005, 2006; Wade and Holt, 2005a). Each stimulus was 589 ms in duration. These speech series members served as categorization targets. Figure 2 shows spectrograms for five series members. Note that creating stimuli using LPC analysis in this way provides the advantage of very natural-sounding tokens, but sound-source separation may be imperfect. The percep-

tual results from previous experiments using these stimuli proved quite clear, so this potential concern is allayed for the present stimulus set.

b. Nonspeech acoustic histories. The tone sequences making up the acoustic histories were created in the same manner as described by Holt (2005), but the distribution variability was explicitly manipulated across conditions. Eight acoustic history distributions defined eight experiment conditions. These distributions were defined by two means, low (1800 Hz) and high (2800 Hz), equivalent to those of Holt (2005). Each of these distribution means was paired with four variability manipulations: 21 unique tone frequencies sampling 1000 Hz in 50-Hz steps; 7 unique tone frequencies sampling 300 Hz in 50-Hz steps (repeated three times each for 21 total tones); 3 unique tone frequencies sampling 100 Hz in 50-Hz steps (repeated seven times each for 21 total tones); or 1 tone at the mean frequency repeated 21 times. For each stimulus, the 21 tones sampling the acoustic history distribution were followed by a 2300-Hz standard tone.

Tones comprising the acoustic histories were synthesized with 16-bit resolution and sampled at 10 kHz using MATLAB (Mathworks, Inc.). Each tone was 70 ms with 5-ms linear onset/offset amplitude ramps. Each tone was followed by 30 ms of silence. As in previous experiments, the order of the tones making up the acoustic history sequence was randomized on a trial-by-trial basis to minimize effects elicited by any particular tone ordering (except, of course, in the case of the 1-tone condition). Consequently, for the 21-, 7-, and 3-tone conditions, each trial was unique; acoustic histories within a condition were distinctive in surface acoustic characteristics, but were statistically consistent with other stimuli drawn from the acoustic history distribution defining the condition. Thus, any influence of the acoustic histories on speech categorization is indicative of listeners' sensitivity to the long-term spectral distribution of the acoustic history and not merely to the simple acoustic characteristics of any particular segment of the acoustic history. Target speech stimuli were digitally down-sampled to 10 kHz from their recording rate of 20 kHz and tones and speech tokens were digitally matched to the rms energy of the /da/ endpoint of the target speech series. A 50-ms silent interval separated the acoustic history plus standard tone sequence and the speech target. Figure 1(a) provides a schematic illustration of stimulus construction and Fig. 1(b) shows a representative stimulus from each condition.

Except for the 1-tone variability conditions, the speech targets were paired with ten unique acoustic histories sampled from the distribution defining the condition to create ten unique repetitions. Stimuli from the conditions with 1-tone variability were repeated ten times each in the experiment. In all, there were 720 stimuli (2 means, 4 variabilities, 9 speech targets, 10 repetitions).

3. Procedure

Seated in individual sound-attenuated booths, listeners categorized the speech target of each stimulus by pressing electronic buttons labeled "ga" and "da." Listeners heard stimuli from each of the eight conditions (720 stimuli) mixed across a single session; each listener responded to stimuli from each condition.

Acoustic presentation was under the control of Tucker Davis Technologies System II hardware; stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified, and presented diotically over linear headphones (Beyer DT-150) at approximately 70 dB SPL(A).

B. Results

Figure 1(c) illustrates listeners' average categorization responses as a function of the acoustic history distribution mean and variability. Figure 1(d) shows the means of the conditions, collapsed across the speech target stimulus series. A 2×4 (mean $\times$ variability) repeated measures ANOVA of the percent "ga" responses averaged across the speech target stimulus series revealed a robust effect of acoustic history distribution mean on speech categorization, $F(1,9)=29.86$, $p<0.0001$, $\eta_p^2=0.768$. Replicating Holt (2005, 2006), listeners more often identified the speech targets as "ga" when the targets were preceded by high mean acoustic histories. However, there was no effect of acoustic history distribution variability on speech categorization, $F(3,27)=1.10$, $p=0.365$, $\eta_p^2=0.109$. There also was no difference in the effect of acoustic history means across variability conditions, as indicated by the lack of an interaction between acoustic history distribution mean and variability, $F<1$. Condition means are presented in Table I.

Although auditory processing shows sensitivity to the mean of a sequence of tones drawn from a spectral distribution, changes in the variability of the distribution did not influence the magnitude of the effect of acoustic histories on speech categorization. It appears that the auditory system does not track the variability of the acoustic distribution as it unfolds or, if it does, it does not play a significant role in producing spectral contrast between contexts and speech targets in this paradigm. It remains possible, of course, that distribution variability may be found to play a significant role in other aspects of speech processing. It has already been implicated, for example, in phonetic category learning in human infants (Maye *et al.*, 2002) and a nonhuman avian species (Kluender *et al.*, 1998). Here, there is considerable evidence for sensitivity to the mean of the distribution, but the variability of the distribution matters remarkably little. The perceptual effect of acoustic histories on speech categorization was statistically indistinguishable whether a single tone at the mean frequency was repeated 21 times or whether 21 unique tones sampling a range of 1000 Hz with the same mean preceded the speech target.

It is interesting to note that this lack of an influence of distribution variability on phonetic context can be differentiated from context effects in visual perception of color. Much like speech, color is perceived relationally; the perceived color at a point in a visual scene depends upon its own characteristics as well as those of the surrounding areas of the scene. Simultaneous color contrast refers to the pattern of perception that arises when changes in the overall brightness or hue of the surround shifts the perceived brightness or hue

TABLE I. Summary statistics for individual conditions.

Experiment	No. of tones overall	No. of unique tone frequencies	Range of tone frequencies (Hz)	Mean "ga" high (SE)	Mean "ga" low (SE)	Main effect acoustic history mean (High versus low)		
						F	$p \le$	n_p^2
1	21	1	1 tone	60.22 (2.89)	45.44 (2.48)	28.32	0.0001	0.759
	21	3	100	60.22 (3.31)	45.67 (2.94)	18.11	0.002	0.668
	21	7	300	60.89 (2.73)	44.78 (2.28)	27.40	0.001	0.753
	21	21	1000	61.67 (2.23)	47.00 (2.25)	28.95	0.0001	0.763
2	21	21 HHH vs. LLL	1000	63.60 (3.07)	38.1 (2.90)	137.98	0.000	0.939
3	21	21 tones	1000	52.17 (2.62)	44.17 (4.05)	19.42	0.002	0.684
	21	21 notched noise	1000	49.33 (2.36)	52.44 (2.30)	5.72	0.04	0.388
4a	3	3	1000	55.67 (1.36)	50.56 (2.40)	4.78	0.05	0.347
	5	5	1000	57.56 (1.15)	53.11 (1.35)	28.8	0.0001	0.762
	6	6	1000	57.33 (1.52)	48.56 (1.75)	26.88	0.001	0.749
	7	9	1000	57.889 (1.68)	48.33 (1.34)	25.96	0.001	0.743
	11	11	1000	59.56 (1.20)	46.44 (1.58)	83.77	0.0001	0.903
	21	21	1000	62.67 (1.51)	46.56 (1.61)	162.43	0.0001	0.947
	26	26	1000	61.44 (1.97)	45.44 (1.69)	46.38	0.0001	0.837
4b	3	3	1000	58.25 (1.73)	58.96 (2.13)	<1	n.s.	0.013
	9	3	1000	60.87 (2.03)	57.78 (1.68)	4.44	0.05	0.254
	15	3	1000	61.75 (2.71)	56.75 (1.93)	4.50	0.05	0.257
	21	3	1000	62.62 (2.39)	54.37 (2.17)	17.15	0.001	0.569
	21	21	1000	64.44 (2.71)	53.97 (1.84)	18.19	0.001	0.583

of the target object in a complementary, contrastive manner (da Vinci, 1492; Chevreul, 1845; Helmholtz, 1866). Brown and MacLeod (1997) report that changing the variance of the colors in a test spot's surround while holding the mean constant induces a contrastive shift in the perceived contrast and saturation of the test spot's color. Thus, as in the auditory case considered in this report, the distribution of visual context influences visual perception. However, in the case of simultaneous visual contrast, distribution variance appears to have a larger role than is observed in the present auditory example.

III. EXPERIMENT 2

Experiment 1 provides further evidence that the auditory system is sensitive to the spectral distribution of context sounds preceding a speech target, particularly the mean. In investigating the processes responsible for such effects, it is of interest whether the entire sequence of the acoustic history is weighted equivalently in its influence on subsequent perception. There are examples of context effects in speech categorization for which temporally adjacent context elicits a more robust effect on speech categorization than does more temporally remote context. For example, whereas the rate of presentation of a precursor sentence influences speech categorization, the segment of the sentence temporally adjacent to the speech target is most influential in target speech syllable identification (e.g., Summerfield, 1975; Newman and Sawusch, 1996).

Experiment 2 addresses this issue in the present paradigm by introducing "meta-statistics" such that the acoustic

history precursors possess both local and global regularities. Previous results demonstrate that higher-frequency acoustic histories shift /ga/-/da/ categorization toward more "ga" responses whereas lower-mean acoustic histories shift categorization toward "da." Holt (2005) showed, as well, that acoustic histories with an intermediate mean (2300 Hz) equivalent to the standard-tone frequency produce a categorization function that is significantly different from those of the high or low mean acoustic histories and lies approximately mid-way between the high and low mean categorization functions.

The present experiment exploits these findings by dividing the 21-tone acoustic history into three sequential 7-tone segments, each with its own local distribution mean. Each of the three segments was synthesized to possess tones sampled from a low-frequency distribution (mean=1800 Hz), a high-frequency distribution (mean=2800 Hz), and an intermediate-frequency distribution (mean=2300 Hz); the order of these three subcomponents of the acoustic history was randomized across conditions, resulting in six distinct orderings each possessing a global mean (across the full 21 tones comprising the entire acoustic history) of 2300 Hz. Whereas in experiment 1 the ordering of the tones making up the acoustic histories was randomized on a trial-by-trial basis to eliminate the potential for local acoustic influences, the present stimulus structure explicitly creates local regularities in the three subcomponents of the acoustic history. A consequence of this is that the 7 tones immediately preceding the speech target varied across conditions, too, making it possible to assess whether, as seems to be the case with rate effects on speech categorization (Summerfield, 1975; Newman and Sawusch, 1996), temporally local information elicits a greater context effect on the speech targets. Most broadly, the question of interest is whether perceptual processing is sensitive to these local statistics or whether the global characteristics of the sequence of tones drive context-dependent speech categorization. To assess this, the effect of these conditions is compared to homogeneous conditions with only global regularities, akin to those of experiment 1 and previous experiments (Holt, 2005, 2006).

A. Methods

1. Participants

Ten volunteers who had not participated in experiments of this kind were recruited from Carnegie Mellon University. They participated for course credit or a small payment. All listeners reported normal hearing and were monolingual native English speakers.

2. Stimuli

Figure 3(a) illustrates the stimulus structure of experiment 2 stimuli. Three conditions were defined by statistically homogeneous acoustic histories, with a global mean but no consistent local regularities. Twenty-one tones sampling a distribution with a high (condition HHH, mean=2800 Hz), middle (condition MMM, mean=2300 Hz), or low (condition LLL, mean=1800 Hz) mean were randomly ordered on a trial-by-trial basis; thus, stimuli in these conditions were
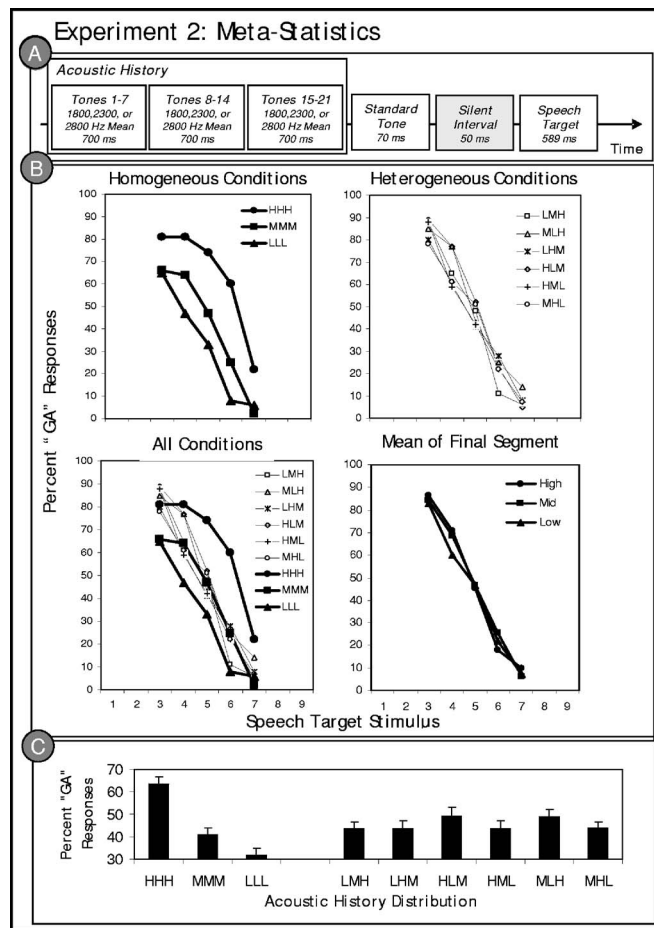
FIG. 3. Stimulus structure and results of experiment 2. (a) Schematic illustration of the experiment 2 stimulus structure; (b) listeners' average "ga" categorization responses to the /ga/ to /da/ series stimuli in the context of preceding acoustic histories. In the top left panel, the thick, solid lines depict speech categorization in the context of the homogeneous acoustic history conditions for which there was only global regularity. In the top right panel, the dashed lines illustrate categorization in the context of the hetergeneous acoustic histories with an intermediate global mean of 2300 Hz and local regularities within 7-tone subelements. The bottom left panel combines the data from the two former graphs. The bottom right panel shows the data as a function of the frequency of the distribution mean of the final subsequence immediately preceding the speech target; and (c) average "ga" categorization responses collapsed across the target stimulus series as a function of the distribution characteristics of the acoustic history precursors.

created identically to those of the most variable distributions in experiment 1, sampling 1000 Hz in 50-Hz steps and possessing only global distribution regularity.

Six conditions possessing local statistical regularity in addition to the global regularity were also created. The acoustic histories of each of these conditions had a global mean of 2300 Hz, equivalent to the middle mean condition of Holt (2005) and the MMM condition of the current experiment. However, unlike previous experiments, these acoustic histories also possessed local regularities in each of three consecutive groups of 7 tones. These local regularities were formed by assigning each of the high, middle, and low distribution means to one of the three 7-tone subsequences of the acoustic history. Thus, averaged across the entire 21 tones, the global mean was always 2300 Hz. However, local regularity existed in each of the subunits of the acoustic history. Exhaustively ordering the high, middle, and low means

across these three subunits yielded six conditions (LMH, LHM, HLM, HML, MLH, MHL). Each of the 7-tone subsequences sampled frequencies around the distribution mean in 50-Hz steps. Therefore, each of the subcomponents of the acoustic history sampled a range of 300 Hz; the global range was 1650–2950 Hz or 1300 Hz. Given the results of experiment 1, this difference in distribution variability from the homogeneous condition was not expected to significantly influence the results. Within a subunit, the 7 tones were randomly ordered on each trial.

To reduce the overall number of observations necessary to test the nine acoustic history conditions, only consonant stimuli 3–7 along the /ga/ to /da/ series were used in experiment 2. Holt (2005) found that this limited range of speech targets, spanning the more ambiguous members of the consonant series, was sufficient to observe effects of acoustic history contexts on speech categorization. Overall, there were 450 stimuli presented in a single experiment session (9 acoustic history conditions, 5 speech target stimuli, 10 repetitions). In all other respects, stimuli were created as in experiment 1.

3. Procedure

The procedure and apparatus were identical to those of the previous experiment. Listeners heard stimuli from each of the nine conditions mixed across a single experiment session. On each trial, listeners responded whether the speech syllable was "ga" or "da."

B. Results

For clarity, the top panels of Fig. 3(b) show separately the categorization functions for the homogeneous conditions possessing only global acoustic regularity and the heterogeneous conditions with local and global regularity. The data are plotted together in the bottom left panel of Fig. 3(b). Condition means can be found in Table I.

A repeated-measures ANOVA was conducted across the nine distribution conditions as a function of the percent "ga" responses averaged across the target stimulus series. This analysis revealed an overall main effect of the acoustic history distribution, $F(8,72)=21.40$, $p<0.0001$, $\eta_p^2=0.704$. Planned Bonferroni-corrected pairwise comparisons revealed that there was a significant difference in listeners' categorization of the consonants preceded by the homogeneous high mean (HHH) and low mean (LLL) conditions, $F(1,9)=137.98$, $p<0.0001$, $\eta_p^2=0.939$. In addition, phonetic categorization was significantly different in the context of the homogeneous middle mean (MMM) condition versus the high (HHH), $F(1,9)=94.02$, $p<0.0001$, $\eta_p^2=0.913$, and the low (LLL), $F(1,9)=23.824$, $p=0.001$, $\eta_p^2=0.726$, homogeneous conditions.

The relationship of speech categorization across the heterogeneous conditions possessing both global and local statistics is of specific interest in the present studies. A repeated-measures ANOVA of the percent "ga" responses across target stimuli for the six heterogeneous conditions with global middle means, but varying local means, revealed no significant main effect of the acoustic histories, $F(5,45)=2.12$, p

$=0.08$, $\eta_p^2=0.191$, although there was a trend in this direction. Investigating this trend more fully among the heterogeneous conditions reveals that only the HLM condition was significantly different from the homogeneous MMM condition (with which heterogeneous conditions shared a global mean). It appears that listeners more often identified consonants following the HLM condition acoustic histories as "ga" than in other conditions with a 2300-Hz global mean, $F(1,9)=30.77$, $p<0.0001$, $\eta_p^2=0.774$. Nonetheless, speech categorization in the context of HLM acoustic histories, albeit distinguishable from the other global intermediate-mean conditions, was significantly different than the HHH condition, $F(1,9)=23.143$, $p=0.001$, $\eta_p^2=0.720$, and the LLL condition, $F(1,9)=50.799$, $p<0.0001$, $\eta_p^2=0.849$. Figure 3(c) illustrates the condition means for each of the nine experiment conditions.

Interpreting these results, it appears that manipulating the local statistics of the acoustic histories has little effect on speech categorization. Despite the presence of local regularities, the global mean dominated the context effect on speech categorization. Irrespective of which local subcomponent to which they belonged, the tones were equal-amplitude sinewaves. Given the overall similarity of the stimuli, grouping at the global level may have been more coherent than that at the local levels, resulting in a bias toward global effects. There thus remains an interesting open question of whether auditory group plays a role in how auditory processing tracks distribution regularity. A straightforward means of beginning to address this possibility in future behavioral work would be to increase the differentiation of the local elements by introducing local changes in amplitude or timbre, for example, to determine whether local regularity exerts more of an influence on defining the distribution when auditory grouping cues are facilitative of it.

Another issue of importance in this experiment was whether local regularities temporally adjacent to the speech targets elicited a greater influence on speech categorization, as has been reported for effects of preceding sentence rate on phonetic categorization (Summerfield, 1975; Newman and Sawusch, 1996). This possibility was investigated with a repeated-measures ANOVA of the percent "ga" responses across the target syllable series as a function of the mean of the final segment of the six heterogeneous conditions. Conditions with a final high mean segment (LMH/MLH) were compared against those possessing a final low mean segment (HML/MHL) and a final middle mean segment (LHM/HLM). The results of this analysis support what is clear from the plot in the bottom right panel of Fig. 3(b). There was no main effect of final segment mean, $F<1$. Phonetic categorization responses could not be predicted from the immediately preceding local regularity alone. Thus, these findings depart from the results of Summerfield (1975) and Newman and Sawusch (1996) for effects of preceding speech on phonetic categorization of temporal contrasts (e.g., /ba/ vs /wa/). Further studies will be needed to determine whether these different patterns of results arise because of differences in temporal versus spectrally based context effects in phonetic categorization, or whether differences in processing speech versus nonspeech contexts may be responsible. Acoustic his-

tories that vary in global temporal characteristics have been found to influence listeners' categorization of following /ba/ vs /wa/ speech targets in a temporally contrastive manner [i.e., slower acoustic histories result in more /ba/ responses, the alternative with a faster formant frequency transition rate (Wade and Holt, 2005c)]. However, it is not yet known whether listeners may be sensitive to local rate regularities in such contexts. Note, as well, that this result is not incompatible with Lotto and Kluender's (1998) finding that adjacent nonspeech tones shift /ga/-/da/ categorization as a function of tone frequency. In the present experiments, the acoustic histories of the present experiment were separated from the speech syllable by a constant mid-frequency standard tone.

IV. EXPERIMENT 3

The first two experiments underscore the relative importance of the global mean of the acoustic history distribution in predicting phonetic context effects. The average spectrum across a somewhat extended temporal window appears to tune processing of incoming acoustic information contrastively with respect to the distribution mean. If this interpretation of the results is correct, then reversing the average energy present in the acoustic history spectra should reverse the influence of context on speech categorization. To be more specific, acoustic histories that *lack* energy at precisely the frequencies where energy is present in the tone-based acoustic histories should produce a context effect on speech categorization in the opposite direction. A complementary spectrum should produce a complementary context effect. This prediction was investigated in experiment 3 by comparing the effects of context for tone acoustic histories and spectral complement acoustic histories created with white noise from which notches were filtered such that acoustic energy existed at all frequencies *except* those sampling the acoustic history distributions.

A. Methods

1. Participants

Ten volunteers who had not participated in experiments of this kind were recruited from Carnegie Mellon University and participated for course credit or a small payment. All listeners reported normal hearing and were monolingual native English speakers.

2. Stimuli

Two sets of acoustic histories were prepared. One set was created from sine-wave tones as described by Holt (2005) and as created for experiments 1 and 2, with high (2800 Hz) and low (1800 Hz) mean distributions sampled by 21 tones with unique frequencies spanning 1000 Hz in 50-Hz steps. As in previous experiments, the order of the acoustic history tones was randomized on a trial-by-trial basis to minimize effects elicited by any particular tone ordering. The final, standard, tone was always 2300 Hz [see Fig. 4(a)]. The speech targets were paired with ten unique acoustic histories sampled from the distribution defining the high and low mean conditions to create ten unique repetitions of each stimulus pairing.
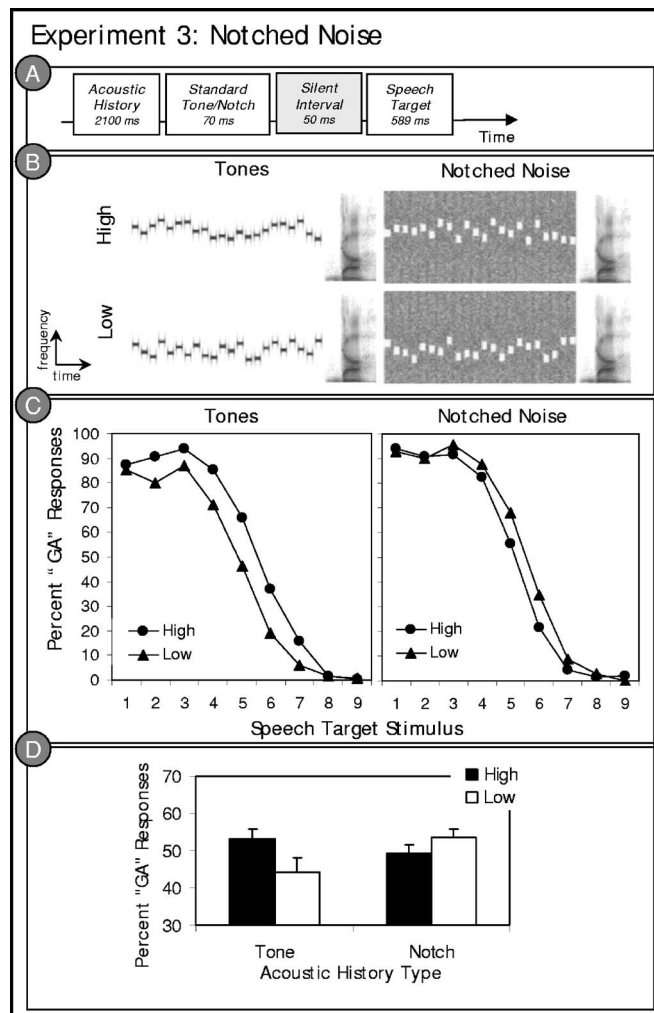
FIG. 4. Stimulus structure and results of experiment 3. (a) Schematic illustration of the experiment 3 stimulus structure; (b) example spectrograms illustrating a high mean and a low mean stimulus token from the notched noise and tone acoustic histories of experiment 3. Note that the tone acoustic histories possess energy where the notched noise acoustic histories lack energy. (c) The left panel illustrates listeners' average "ga" categorization responses to the /ga/ to /da/ stimulus series as a function of the tone acoustic history contexts. The right panel shows the same listeners' responses to the same speech stimuli preceded by notched noise acoustic history contexts. (d) Average "ga" categorization responses collapsed across the target stimulus series as a function of the distribution characteristics of the acoustic history precursors.

A matched set of notched noise acoustic histories was also created. Stimulus for stimulus, the frequencies and orderings of the tone acoustic histories were matched. However, the spectra for the notched noise acoustic histories were created as spectral complements to the tone acoustic histories. This was achieved by beginning with a white noise stimulus possessing spectral energy across all frequencies. From this 2200-ms white noise stimulus, notches (70-ms, 100-Hz-bandwidth spectral regions with acoustic energy greatly reduced) were produced by digitally band-stop filtering the white noise in sequential time slices (Matlab, Mathworks; sixth-order elliptical filter, with 3-dB peak-to-peak ripple and a minimum stop-band attenuation of 50 dB). Notch center frequencies were equivalent to tone frequencies in the tone acoustic histories. As a result, these notched-noise

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Lori L. Holt: Characteristics of context dependence   2809

stimuli were spectral complements to the tone acoustic histories in that they lacked acoustic energy precisely where the tone acoustic histories possessed energy. In other ways, stimulus creation was as described for the previous experiments. Example stimuli are pictured in Fig. 4(b). In all, there were 360 stimuli (2 means, 2 types of acoustic history, 9 speech target stimuli, 10 repetitions).

3. Procedure

Seated in individual sound-attenuated booths, listeners categorized the speech target of each stimulus by pressing electronic buttons labeled "ga" and "da." Listeners completed two blocks in a single session; the order of the blocks was counterbalanced. In one block, the notched noise contexts preceded the speech targets. In the other block, participants heard the same speech targets preceded by tone acoustic histories. Thus, each listener responded to stimuli from each condition.

B. Results

Figure 4(c) depicts listeners' average categorization responses as a function of the acoustic history type (tone versus notched-noise) and acoustic history distribution mean (high versus low). Figure 4(d) illustrates the condition means, collapsed across the speech target series stimuli. Condition means can be found in Table I. A repeated measures ANOVA of the percent "ga" responses across the speech target series reveals what the graphs illustrate: both the tone acoustic histories, $F(1,9)=19.46$, $p=0.002$, $\eta_p^2=0.684$, and the notched noise acoustic histories, $F(1,9)=5.72$, $p=0.04$, $\eta_p^2=0.388$, had a significant effect on speech categorization. Of note, however, is the direction of these context effects. A direct prediction of spectral contrast is that notched-noise spectral complements of tonal acoustic histories (created such that where there was energy in the tone acoustic histories there as a corresponding lack of energy in the notched noise) should produce the *opposite* effect of context on speech categorization. This prediction is cleanly supported by the present data. As reported in previous studies (Holt, 2005, 2006) and experiments 1 and 2, the tone acoustic histories shifted speech categorization contrastively. Speech targets preceded by acoustic histories sampled from a higher-frequency spectral distribution were more often categorized as "ga," the lower-frequency alternative, than the same stimuli preceded by tone acoustic histories sampling a lower-frequency spectral distribution. The notched noise acoustic histories produced results that similarly follow from the distributions of spectral energy. The notched noise acoustic histories labeled "high" had notches modeling the frequencies of the "high" tone acoustic histories (and lacked acoustic energy around the high mean, 2800 Hz). As a result, they possessed relatively *less* high-frequency energy than the "low" notched noise stimuli. Listeners' categorization responses followed this spectral reversal. Listeners more often categorized consonants preceded by "low" notched noise acoustic histories as "ga" relative to the same consonants preceded by "high" notched noise acoustic histories. Again, greater high-frequency energy in the precursor stimulus led

to speech categorization functions contrastively shifted toward lower-frequency spectral energy and vice versa. These results therefore substantiate the claim that the average spectrum of acoustic energy preceding a speech target has a significant spectrally contrastive influence on speech categorization.

A further analysis was conducted to examine whether the size of the context effect produced by the notched noise versus tone acoustic histories differed. Participants' percent "ga" categorization responses were submitted to a 2×2 (context type $\times$ acoustic history mean) repeated measures ANOVA of the percent "ga" responses across the speech target series. The analysis revealed that there was no main effect of whether the acoustic history was created from tones or notched noise, $F<1$, indicating no overall shift in speech categorization across context types. As would be expected from the analyses above, there was a significant influence of the acoustic history distribution mean, $F(1,9)=28.38$, $p=0.0005$, $\eta_p^2=0.759$. Of note, there was no context type by acoustic history distribution mean interaction, $F(1,9)=2.92$, $p=0.122$, $\eta_p^2=0.245$. Context effects produced by the notched noise versus the tone acoustic histories were statistically equivalent, indicating that the size of the context effects produced by the tone and notched-noise acoustic histories were statistically equivalent in magnitude, albeit complementary. Caution, however, should be taken in interpreting the importance of this null interaction. The principal finding of experiment 3 is that spectrally complementary acoustic histories produce a complementary effect on phonetic categorization. The effect is predicted from the average spectrum of the preceding nonspeech context.

V. EXPERIMENT 4

From the results so far, we can conclude that the mean matters in auditory context effects. In the face of considerable variability or no variability, with local or only global regularity, and with tones or notched noise, the global mean spectra of the acoustic history distribution predicts the effect on speech categorization. Moreover, the effects are "statistical" in the sense that acoustic characteristics that vary trial-by-trial do not drive them. Rather, they are related to *distributional* characteristics of the context. Distributions, however, must be sampled; the greater the number of events drawn from a distribution, the better the estimate of that underlying distribution parameter such as the mean.

Does the auditory system treat distributional information in this way? Are the distributional context effects elicited by acoustic histories more robust when the distributions are better sampled with more events across time or with finer sampling of frequency within the spectral range? Table I illustrates that the experiments reported thus far have investigated acoustic histories sampled by an equivalent number of tones. In this scheme, the duration of the acoustic history has been constant as has been the number of tones sampling the distribution. To return to the original aim of the present studies to investigate the bounds of sensitivity to distribution regularity in auditory processing, experiment 4 investigates whether sampling the acoustic history with
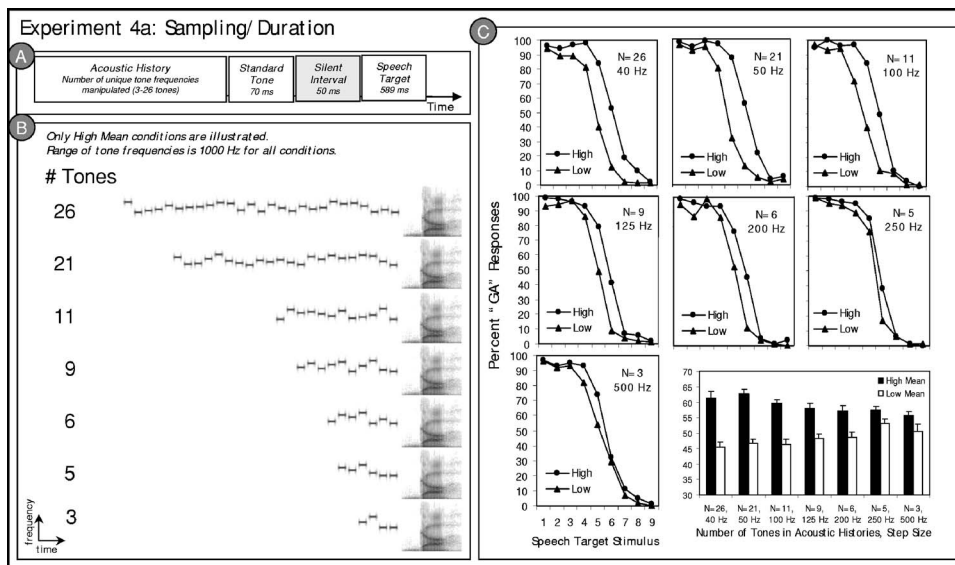
FIG. 5. Stimulus structure and results of Experiment 4a. (a) Schematic illustration of the experiment 4a stimulus structure; (b) example spectrograms illustrating a high mean stimulus token from each experiment 4a condition. Low mean conditions (not illustrated) follow the same pattern across conditions. (c) Listeners' average "ga" categorization responses to the /ga/ to /da/ series stimuli in the context of preceding high and low mean acoustic histories are plotted for each of the seven conditions of Experiment 4a. The bar graph in the lower right corner illustrates the same data collapsed across speech target stimulus.

greater density across frequency (experiment 4a) or more events across time (experiment 4b) modulates the influence of the distribution mean on phonetic categorization. In experiment 4a, the number of tones composing the acoustic histories is manipulated while the range of frequencies sampled is constant at 1000 Hz. As a result, the density with which the distribution is sampled varies (as reflected by the step-size between tones comprising the distribution). If three tones with unique frequencies sample a 1000-Hz range, then the sampling frequency of these tones must be greater (500 Hz) than the sampling frequency of 26 unique tones that sample a 1000-Hz range (40 Hz) when distribution mean is held constant. Due to the multiple dimensions defining the acoustic history distribution in the spectral and temporal domains, this scheme necessarily covaried acoustic history duration and frequency sampling density, thus confounding firm conclusions as to the source of any observed effects.

Therefore, a second study was run with additional controls. Experiment 4b examines speech categorization in the context of acoustic histories that vary in the sampling of events across time by varying the number of tones comprising the acoustic histories (from 3 to 21 tones), with the sampling density across frequency held constant. Jointly, these experiments examine whether the distributional effects of auditory context observed on phonetic categorization are subject to manipulations of distribution sampling in the frequency and time domains.

A. Method

1. Participants

Ten volunteers who had not participated in experiments of this kind were recruited from Carnegie Mellon University and participated for course credit or a small payment. All listeners reported normal hearing and were monolingual native English speakers.

2. Stimuli

a. Experiment 4a. Stimuli were created in the manner of experiment 1, with the overall structure following the pattern schematized in Fig. 5. In experiment 4a the number of tones in the sequence composing an acoustic history was manipulated across seven conditions such that high and low spectral distributions were sampled with 3, 5, 6, 9, 11, 21, or 26 tones. The 21-tone condition was created in the same way as acoustic histories of experiments 1–3, with tones sampled at 50-Hz intervals. The range of variability of tones was held constant at 1000 Hz across conditions, so the frequency distance between tones composing an acoustic history varied (3 tones, 500 Hz; 5 tones, 250 Hz; 6 tones, 200 Hz; 9 tones, 125 Hz; 11 tones, 100 Hz; 21 tones, 50 Hz; 26 tones, 40 Hz). Table I provides a summary. The 2300-Hz standard tone served as the final tone in each sequence. Each stimulus was repeated ten times for a total of 1260 stimuli (7 sampling density conditions × 2 acoustic history means × 9 speech target stimuli × 10 repetitions) in the experiment.

Overall, the present stimuli differed from those of the previous experiments in three ways: (1) the number of tones in the acoustic history varied from 3 to 26; (2) tone duration was held constant, so overall duration of the acoustic histories varied with the number of tones (from 300 to 2600 Hz); and (3) the overall range of frequencies sampled was held constant at 1000 Hz, so the density of frequency sampling across this 1000 Hz varied with the number of tones composing the acoustic history (40–500 Hz, from 26 to 3 tones).

b. Experiment 4b. Holding the overall range of the distribution and the tone duration constant in experiment 4a meant that the density of frequency sampling across the 1000-Hz range and the overall duration of the acoustic histories varied simultaneously. Therefore, it is not possible to know definitively the source of any observed effects of the acoustic histories of experiment 4a. To address this concern, an additional set of acoustic histories was created for experiment 4b. In these stimuli, the distribution sampling density was held constant at 500 Hz; tones in the experiment were sampled from the distributions in 500-Hz steps. Three unique tones with frequencies sampling the low mean (1300, 1800, 2300 Hz) distribution or the high mean (2300, 2800, 3300 Hz) distribution served as the basis for each of the
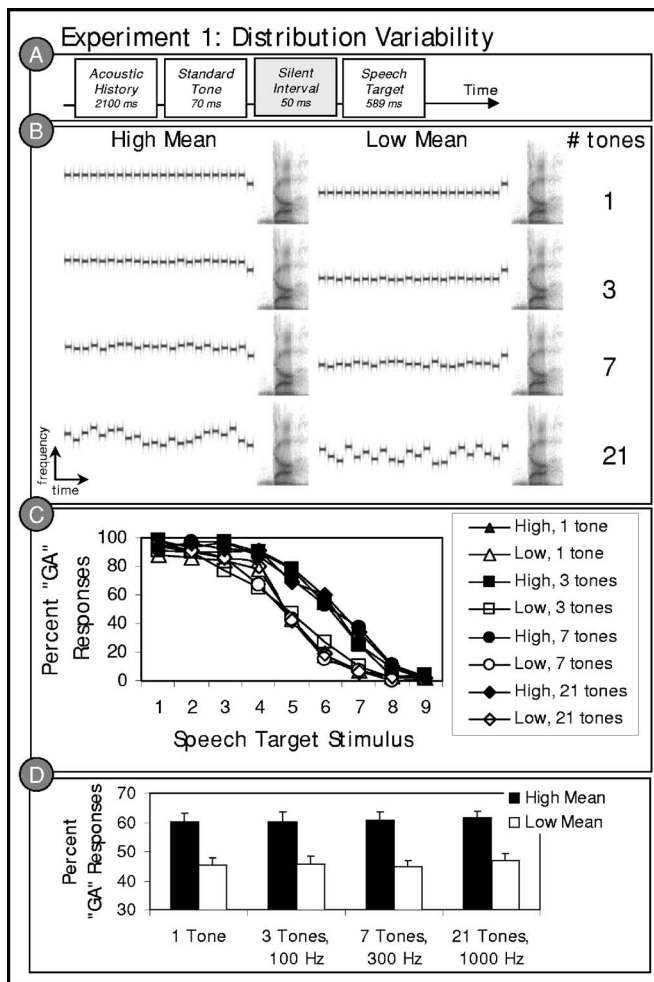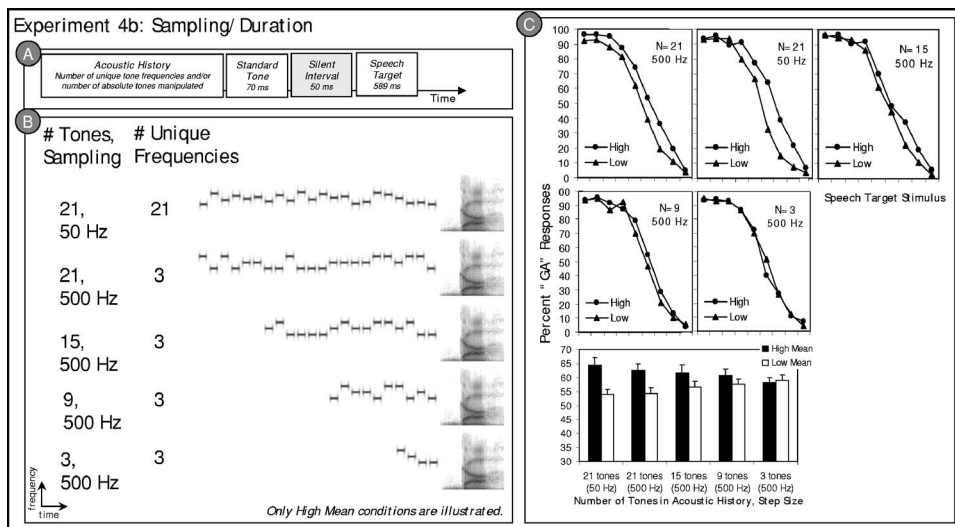
FIG. 6. Stimulus structure and results of Experiment 4b. (a) Schematic illustration of the experiment 4b stimulus structure; (b) example spectrograms illustrating a high mean stimulus token from each experiment 4b condition. Low mean conditions (not illustrated) follow the same pattern across conditions. (c) Listeners' average "ga" categorization responses to the /ga/ to /da/ series stimuli in the context of preceding high and low mean acoustic histories are plotted for each of the five conditions of experiment 4b. The bar graph in the lower right corner illustrates the same data collapsed across speech target stimulus.

acoustic histories. These three tones were repeated one, three, five, or seven times to create acoustic histories composed of 3, 9, 15, or 21 tones. The order of the tones was randomized on a trial-by-trial basis. There were 900 total stimuli (5 conditions × 2 acoustic history means × 9 speech target stimuli × 10 repetitions). In other ways, stimuli were created as described for experiment 1; representative stimuli are shown in Fig. 6.

The experiment 4b stimuli had a constant sampling of frequency (1000 Hz in 500-Hz steps), but the duration (300–2600 ms) varied with the number of tones composing the acoustic histories (3 to 26 tones). For comparison and as a control condition, an additional pair of high and low mean acoustic history conditions was created with 21 tones sampling 1000 Hz in 50-Hz steps, as in experiments 1–3. Overall, comparing the results of experiments 4a and 4b, it is possible to begin to understand whether distribution regularity is related to the density with which the acoustic spectrum is sampled or the absolute duration of the acoustic history.

3. Procedure

Seated in individual sound-attenuated booths, listeners categorized the speech target of each stimulus by pressing electronic buttons labeled "ga" and "da." Listeners responded to stimuli from each condition mixed across an experimental session on each of two consecutive days.

B. Results

The results of experiment 4a are plotted in Fig. 5 and were investigated with a 7 × 2 (number of tones × acoustic history mean) repeated measures ANOVA of the percent "ga" responses averaged across the target speech series. As was the case in experiments 1–3, there was a robust main effect of acoustic history distribution mean on speech categorization, $F(1,9)=90.04$, $p<0.0001$, $\eta_p^2=0.909$, with listeners responding "ga" more often in the context of preceding high mean acoustic histories. There was no main effect of the number of tones in the acoustic histories, $F(6,54)=1.89$, $p=0.10$, $\eta_p^2=0.174$, indicating that there was no overall categorization shift as a function of the number of tones composing the acoustic history. Of interest, there was an interac-

tion, $F(6,54)=10.44$, $p<0.0001$, $\eta_p^2=0.537$, indicating that the size of the context effect elicited by the acoustic history mean varied as a function of the number of tones composing the acoustic history. The magnitude of the effect of the acoustic history mean on speech categorization decreased with the number of tones composing the distribution. Table I provides condition means.

Thus, it appears that the effect of the preceding spectral distribution on speech categorization may be modulated by the acoustic history duration or the number of events composing it. However, to satisfy the multiple constraints on stimuli in experiment 4, the number of tones composing the acoustic histories across conditions necessarily covaried with the sampling of the range of spectral distribution. Therefore, it is impossible to know the source of the interaction of acoustic history distribution mean with the number of tones composing the acoustic histories. The number of acoustic events, the overall acoustic history duration, and the density with which the acoustic history was sampled varied across the conditions so the source of the interaction cannot be inferred from experiment 4a alone.

To investigate this further, experiment 4b examined the effect of acoustic histories for which the number of tones varied across conditions but the mean of the distribution and density with which the acoustic histories were sampled were held constant. Acoustic histories with 3, 9, 15, or 21 tones sampled at 500-Hz intervals and a 21-tone acoustic history sampled at 50 Hz, as in previous experiments, served as contexts. The interesting comparisons come among the perceptual consequences of the four conditions sampled at 500 Hz, for which the number of tones varied while sampling interval was held constant, and between the two 21-tone conditions for which sampling frequency varied while the number of tones was constant.

A 4 × 2 (sampling interval × acoustic history mean) repeated measures ANOVA of the percent "ga" responses across the speech target series comparing the conditions for which sampling density was constant revealed no overall effect of the number of tones, $F<1$, but a significant effect of the acoustic history distribution mean, $F(1,13)=6.23$, $p=0.02$, $\eta_p^2=0.324$. The interaction between the number of

tones and the acoustic history distribution mean was also significant, $F(3,39)=8.53$, $p=0.0002$, $\eta_p^2=0.396$, indicating that the size of the context effect varied with the number of tones composing the acoustic history. With fewer tones, the effect of the distribution means was smaller. Table I provides condition means. Figure 6 plots the results.

An additional 2×2 (sampling density $\times$ acoustic history mean) repeated measures ANOVA of the percent "ga" responses across the target series, comparing just the conditions for which sampling interval varied (50 vs 500 Hz) while the number of tones (21) remained constant, revealed no significant overall effect of condition, $F<1$, but the expected effect of acoustic history distribution mean, $F(1,13)=20.157$, $p=0.001$, $\eta_p^2=0.608$. Of special interest, there was no interaction of acoustic history distribution mean with the sampling density of the tones, $F(1,13)=1.896$, $p=0.192$, $\eta_p^2=0.127$. Thus, the context effect was the same magnitude for 21-tone acoustic histories with high and low distribution means, whether the 21 tones sampled 21 unique frequencies in 50-Hz steps across 1000 Hz or whether they sampled just three frequencies in 500-Hz steps across the 1000-Hz range. When the acoustic history distribution mean and number of tones was held constant, there was no effect of sampling density.

The change in the magnitude of the context effects observed for experiment 4a therefore appears to be the result of the distribution mean and the duration of the acoustic history, or the number of events composing the acoustic history, rather than the density with which the distribution was sampled in the frequency domain. The longer acoustic histories produced larger effects of context on speech categorization, the direction of which was predicted by the distribution mean for all conditions. Again, as in the previous experiments, the mean of the distribution plays the dominant role in driving the effects of context observed on speech categorization.

VI. GENERAL DISCUSSION

Like perception in other modalities, auditory perception is relational in the sense that context is critical in interpreting information for object identity. From the lowest-level cochlear interactions to higher-order cortical processing, the auditory neural code is significantly affected by context sounds. Nonetheless, not all characteristics of context are effective in influencing perception. As a result, understanding the characteristics of context that have a perceptual effect can be informative of the types of information that interact in perceptual processing. The present studies find that nonlinguistic acoustic sounds as simple as tones and notched white noise are sufficient to influence speech categorization. This implicates rather general perceptual processes in the categorization of speech (Diehl *et al.*, 2004; Holt and Lotto, 2002; Lotto and Kluender, 1998).

Holt (2005) introduced an "acoustic history" paradigm for studying the effects of statistical distributions of nonspeech context sounds on speech perception and demonstrated that manipulating the spectral mean of a distribution sampled by a sequence of tones of varying frequencies pro-

duces robust spectrally contrastive context effects on speech categorization. The present series of experiments exploited the acoustic history paradigm to zero in on the statistical distribution characteristics that influence online speech categorization. Experiment 1 found that whereas the mean of the acoustic history distributions serves to shift speech categorization contrastively, the variability of the distributions has little effect. Whether the acoustic history distribution was sampled by 21 tones with unique frequencies spanning a 1000-Hz range or a single tone at the mean frequency of that range repeated 21 times, the effect on speech categorization was robust and statistically equivalent. Experiment 2 demonstrated that local statistical regularity existing within the global regularity did not have much of an effect on speech categorization. Listeners' speech target identification was influenced by the global mean of the acoustic history, without significant influence of the local regularity. Experiment 3 provided additional evidence that the average spectrum of the acoustic histories drives the observed context effects; spectral complements of the acoustic histories produced an opposite effect on speech categorization. The spectrum thus predicts the direction of the context effects. Finally, experiment 4 highlighted how distribution sampling influences the effect of the spectral distribution mean on speech perception. The number of tones composing the acoustic history was influential; as the number of tones and the overall duration of the acoustic history decreased, so too did the effect of the acoustic history mean on speech categorization. The sampling density of the distribution in the frequency domain was not essential, however. Whether the frequency range was sampled with tones differing by 500 or 50 Hz, the effect of the acoustic histories on speech categorization was the same so long as the number of tones and duration of the acoustic history were constant; longer acoustic histories produced more robust context effects. Moreover, the context effects observed across all of the experiments strongly support the claim by Holt (2005) that the mean of a distribution of context sounds influences speech categorization. Table I provides a verification of the robust nature of the influence of acoustic history contexts on speech categorization. Across conditions and experiments, this finding was replicated 18 times.

On the whole, the present experiments provide converging evidence for auditory sensitivity to spectral distributions of sound and of perceptual consequences of this sensitivity in processing other sounds. There are three primary characteristics of these data that must be understood. The first is that nonspeech contexts are effective in shifting speech categorization at all. The second is the contrastive nature of the results, such that lower frequency contexts shift speech categorization toward higher-frequency alternatives and higher-frequency contexts shift speech categorization toward lower-frequency alternatives. These first two characteristics are supported thoroughly by other studies of nonspeech contexts on speech categorization. The context effects produced by the acoustic histories mirror the contrastive effects reported for single, temporally adjacent nonspeech contexts (Holt, 1999; Holt *et al.*, 2000; Lotto and Kluender, 1998), although with longer time courses (Holt, 2005). The third characteris-

tic of importance was hinted at in the studies of Holt (2005), but is firmly supported by the present studies: the mean matters. The influence of the distributionally defined nonspeech acoustic history contexts was driven by the mean of the distribution across experiments. At a broad, algorithmic level a general auditory perceptual mechanism that accentuates change such that, in response to regularity in the acoustic input, the system responds best to novelty, would accommodate these three primary characteristics of the results.

A. Stimulus-specific adaptation and spectral contrast

Neuronal adaptation, the decline over time of neuron responses to sensory stimuli, serves this algorithmic requirement. It is a ubiquitous characteristic of neural processing in sensory areas and, in the auditory system, neuronal adaptation exists at multiple levels. Auditory nerve neurons are able to follow very fast acoustic presentation rates of up to hundreds of presentations per second (Greenwood and Joris, 1996), but at the level of the auditory thalamus (medial geniculate body, MGB) neurons adapt at rates of tens of presentations per second (Creutzfeldt et al., 1980). Cortical neurons in primary auditory cortex (A1) adapt even when stimulus presentation is at a rate of only a few stimuli per second (Condon and Weinberger, 1991; Creutzfeldt et al., 1980; Miller et al., 2002). Thus, neurons show somewhat weak, fast adaptation at the periphery, but adaptation becomes progressively stronger, with longer time constants along the ascending auditory pathway.

Adaptation is often thought of simply as a fatigue of the neural response and, in fact, it can be manifested in this way (Carandini, 2000). In past work examining the contrastive nature of nonspeech context effects on speech, we had hypothesized that peripheral adaptation of this sort might be responsible for the effects of single temporally adjacent speech and nonspeech context stimuli on speech categorization (Holt et al., 2000). Subsequent research, however, revealed context effects of nonspeech even when the nonspeech signals were presented to the ear opposite that of the speech target and across silent intervals incompatible with solely peripheral processing (Lotto et al., 2003; Holt and Lotto, 2002).

A fairly recent series of studies has provided strong evidence that the auditory system, like the visual system (Dragoi et al., 2000; Movshon and Lennie, 1979; Saul and Cynader, 1989), exhibits another form of adaptation—known as *stimulus-specific adaptation* (SSA)—that may better account for spectrally contrastive context effects. This stimulus-specific adaptation has been shown to occur with long adaptation sequences (Condon and Weinberger, 1991; Dragoi et al., 2000; Movshon and Lennie, 1979; Saul and Cynader, 1989) and also with pairs of stimuli (Malone and Semple, 2001; Malone et al., 2002; Müller et al., 1999). Thus, SSA can occur rapidly (on the time scale of a single stimulus) or evolve more slowly over multiple stimulus presentations. Visual neurons, for example, adapt to the statistical distribution of input stimuli. This pattern of SSA has been shown to maximize information transmission in the visual network and

is closely related to principles of optimal neural coding (Fairhall et al., 2001).

Ulanovsky and colleagues (Ulanovsky et al., 2003, 2004) have recently described SSA in primary auditory cortex (A1) and have suggested that a central role of A1 may be to respond to regularity in the long-term statistics of auditory signals by depressing neuron response to regularity in the acoustic input and, relatedly, to detect auditory novelty with enhanced neuronal response. At the algorithmic level, SSA is perhaps better suited than adaptation-by-fatigue to relate to context-dependent auditory processing of the sort investigated here. Ulanovsky et al. have provided explicit evidence that SSA creates *enhancement* of neural responses to rare stimuli, and thus to acoustic change and the *stimulus specificity* of the neural adaptation and enhancement. The effects are therefore sensitive to distributions of energy evolving across time and elicit adaptation and the consequent enhancement that would be expected to produce contrastive patterns of perception like those observed in the present results.

The experimental paradigm of Ulanovsky et al. (2003) can be considered to be a single-neuron version of mismatch negativity (Näätänen et al., 1978), an important auditory event-related potential (ERP) whereby sequences of common acoustic stimuli are presented with rare deviant stimuli occasionally embedded in an "oddball" design. The MMN is the difference in the ERP for the deviant versus the standard, with an enhanced response to the deviant. The MMN can reflect deviation in stimple acoustic properties like absolute frequency or duration, or complex regularities in acoustic input that evolve across time. Given that the perceptual system must have some accounting of what is standard in order to respond more robustly to a deviant, many have interpreted the existence of the MMN as indicative of auditory sensory memory (see Näätänen et al., 2001). Employing a standard/deviant "oddball" stimulus presentation paradigm in studying single-neuron responses in A1, Ulanovsky et al. report an enhanced response of individual auditory cortical neurons to acoustic novelty. Specifically, a particular tone frequency serving as an infrequent deviant in a series of tones is responded to more robustly by A1 neurons than is the same tone frequency when it serves as a frequent, repeating standard. These data provide evidence that primary auditory cortex tracks statistical regularity in the input signal across rather extended temporal windows and modulates its responsiveness in reaction to this regularity. The relative adaptation to common input characteristics coupled with the enhanced response to novelty serves to exaggerate acoustic change in the neural code. The coding strategy of the auditory system is thus matched to the stimulus statistics over time (see also Fairhall et al., 2001).

Of course, these physiological studies were not conducted with the acoustic history paradigm of the present experiments in mind and so direct comparisons should be made with much caution. However, these studies do demonstrate the sensitivity of auditory cortex to distribution characteristics of acoustic context that unfolds across time. With this caveat in mind, there is a way in which SSA may account for the robust effect of distribution mean in influencing phonetic

Lori L. Holt: Characteristics of context dependence

categorization in the present experiments. In addition to studying A1 neuron response to distributions of stimuli that varied in their regularity, Ulanovsky *et al.* (2004) examined the response of A1 neurons to equally probable, equal-amplitude tones with 20 different frequencies. This stimulus ensemble was thus quite similar to the acoustic histories of the present experiments. Ulanovsky *et al.* report that the responses of the A1 neurons to frequencies at the center of the stimulus frequency range adapted the most and there was relative *enhancement* of responses at the eccentric frequencies furthest away from the center of the frequency range. This response pattern created a U-shape in the neural tuning curves, with maximal adaptation at the central frequencies and enhancement at the edges. This response pattern emerged due to relatively simple properties of SSA. Adaptation strength is negatively correlated with the frequency difference between the present stimulus and that of the stimulus from the preceding trial (Brosch and Schreiner, 1997; Ulanovsky *et al.*, 2004) and, on average, central frequencies have smaller frequency differences from the preceding trials compared to eccentric frequencies. As a result, adaptation is greatest for central frequencies and, significantly, the response of the neuron to the eccentric frequencies is enhanced. This pattern of neural coding meets the algorithmic requirements necessary to explain the importance of the distribution mean in predicting the patterns of perception observed in the present experiments. The possibility that such a mechanism is consistent with the robust perceptual effects of distribution regularity on phonetic categorization is intriguing, but should not be over-interpreted without further explicit behavioral and physiological investigations.

B. Implications for speech processing

It is rather unexpected that a simple sequence of tones something like a tune should influence speech perception. However, considered in light of the natural soundscape in which we have evolved, the results of the present studies need not be entirely surprising. Sounds in the natural world rarely appear in isolation; rather, they are intermixed with common and rare acoustic events that change on multiple time scales (Nelken *et al.*, 1999). Added to this complexity is the fact that sounds exist for mere moments. Thus, there are significant demands from the environment for fast, online adaptation to regularity as well as sensitivity to longer-term regularity in the acoustic input. Perceptual processing does not provide a passive, veridical representation of the acoustic environment; instead it provides a means of rapidly extracting information relevant to behavior in the present circumstances. So, too, when the acoustic signal is speech. From this perspective, the context in which a signal occurs is critical to its perception. Sensitivity to regularity and concomitant detection of change appears to be characteristic of auditory processing and may provide us with insights about behaviorally relevant questions in speech processing.

As an example, it has long been observed that listeners perceive vowel categories dependent on the spectral characteristics of a precursor sentence (Ladefoged and Broadbent, 1957) such that they appear to normalize perception for different talkers or accents. Many models suggest a speech-specific translation from the surface acoustic variability to a more canonical representation [see Repp (1982) for a review]. The nonlinguistic nature of the context effects elicited by the acoustic histories, coupled with their sentence-level time course, suggests the possibility such "normalization" may arise when spectral regularities in the acoustic environment produce contrastive effects on auditory processing and, concomitantly, speech categorization. The present results specifically suggest that sensitivity to the average spectrum of a particular speaker's speech may help to resolve ambiguities in speech categorization later in an utterance through rather general auditory perceptual mechanisms not specific to speech, but nonetheless operating at a rather high level of auditory processing sensitive to statistical regularities unfolding across sentence-length time courses. This, of course, does not deny the importance of longer-term learning and generalization to familiar voices that is known to occur in speech processing (e.g., Johnson and Mullennix, 1997). However, it provides a means of reconsidering one of the classic issues in speech perception: how do listeners maintain perceptual constancy despite the considerable acoustic variability present in the speech signal? The "lack of invariance" between acoustics and speech categories has long been taken as evidence of the special status of speech perception relative to auditory perception in general (see Repp, 1982). Considering the effects of nonspeech contexts on speech categorization and the general perceptual mechanisms that might relate to them casts doubt on whether this supposition is necessary. Rather, the present data would argue, it is necessary to consider the fact that the problem of perceptual constancy in the presence of variable contexts is a very general problem faced by all perceptual systems, whatever the signal. Given the strong evidence of nonspeech context effects on speech categorization (Coady *et al.*, 2003; Fowler *et al.*, 2000; Holt, 1999; Holt and Kluender, 2000; Holt *et al.*, 2000; Holt, 2005; Lotto *et al.*, 2003; Lotto and Kluender, 1998) and the demonstration that a nonhuman species exhibits "phonetic" context effects (Lotto *et al.*, 1997), it appears that speech communication makes use of general characteristics of perceptual processing to maintain perceptual constancy in the face of variability. The present results highlight that this may be accomplished by an adaptive tendency of the auditory perceptual system to seek regularity in acoustic input and exaggerate change relative to the regularity. Specifically, in the case of distributional regularity in the acoustic environment, the average frequency across time appears to be influential in auditory processing.

ACKNOWLEDGMENTS

[1]A recent investigation suggests that such effects also extend to temporal characteristics of the acoustic environment (Wade and Holt, 2005c).

Brosch, M., and Schreiner, C. (**1997**). "Time course of forward masking tuning curves in cat primary auditory cortex," J. Neurophysiol. **77**, 923–943.

Brown, R., and MacLeod, D. (**1997**). "Color appearance depends on the variance of surround colors," Curr. Biol. **7**, 844–849.

Carandini, M. (**2000**). "Visual cortex: Fatigue and adaptation," Curr. Biol. **10**, R605–R607.

Chevreul, M. E. (**1845**). *The Laws of Contrast of Colours and Their Application to the Arts* (Routledge, London).

Coady, J. A., Kluender, K. R., and Rhode, W. S. (**2003**). "Effects of contrast between onsets of speech and other complex spectra," J. Acoust. Soc. Am. **114**, 2225–2235.

Condon, C., and Weinberger, N. (**1991**). "Habituation produces frequency-specific plasticity of receptive fields in the auditory cortex," Behav. Neurosci. **105**, 416–430.

Creutzfeldt, O., Hellwag, F., and Schreiner, C. (**1980**). "Thalamocortical transformation of responses to complex auditory stimuli," Exp. Brain Res. **39**, 87–104.

Darwin, C. J., McKeown, J. D., and Kirby, D. (**1989**). "Perceptual compensation for transmission channel and speaker effects on vowel quality," Speech Commun. **8**, 221–234.

da Vinci, L. (**1492**). *Treatise on Painting* (Princeton U.P., Princeton, NJ).

Diehl, R., Lotto, A., and Holt, L. L. (**2004**). "Speech perception," Annu. Rev. Psychol. **55**, 149–179.

Dragoi, V., Sharma, J., and Sur, M. (**2000**). "Adaptation-induced plasticity of orientation tuning in adult visual cortex," Neuron **28**, 287–298.

Fairhall, A. L., Lewen, G. D., Bialek, W., and de Ruyter van Steveninck, R. R. (**2001**). "Efficiency and ambiguity in an adaptive neural code," Nature (London) **412**, 787–792.

Fowler, C. A., Brown, J. M., and Mann, V. A. (**2000**). "Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans," J. Exp. Psychol. Hum. Percept. Perform. **26**, 877–888.

Francis, A. L., Ciocca, V., King Yu Wong, N., Ho Yin Leung, W., and Cheuk Yan Chu, P. (**2006**). "Extrinsic context affects perceptual normalization of lexical tone," J. Acoust. Soc. Am. **119**, 1712–1726.

Gow, D. W. (**2003**). "Feature parsing: Feature cue mapping in spoken word recognition," Percept. Psychophys. **65**, 575–590.

Greenwood, D. D., and Joris, P. X. (**1996**). "Mechanical and 'temporal' filtering as codeterminants of the response by cat primary fibers to amplitude-modulated signals," J. Acoust. Soc. Am. **99**, 1029–1039.

Hauser, M. D., Newport, E. L., and Aslin, R. N. (**2001**). "Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins," Cognition **75**, 1–12.

Helmholtz, H. (**1866**). *Physiological Optics* (Optical Society of America, Rochester, NY).

Holt, L. L. (**1999**). "Auditory constraints on speech perception: An examination of spectral contrast," Diss. Abstr. Int., B **61**, 556.

Holt, L. L. (**2005**). "Temporally non-adjacent non-linguistic sounds affect speech categorization," Psychol. Sci. **16**, 305–312.

Holt, L. L. (**2006**). "Speech categorization in context: joint effects of non-speech and speech precursors," J. Acoust. Soc. Am. **119**, 4016–4026.

Holt, L. L., and Kluender, K. R. (**2000**). "General auditory processes contribute to perceptual accommodation of coarticulation," Phonetica **57**, 170–180.

Holt, L. L., and Lotto, A. (**2002**). "Behavioral examinations of the level of auditory processing of speech context effects," Hear. Res. **167**, 156–169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.

Johnson, K. J., and Mullennix, J. W. (**1997**). *Talker Variability in Speech Processing* (Academic, San Diego).

Kiefte, M. J., and Kluender, K. R. (**2001**). "Spectral tilt versus formant frequency in static and dynamic vowels," J. Acoust. Soc. Am. **109**, 2294–2295.

Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. B. (**1998**). "Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models," J. Acoust. Soc. Am. **104**, 3568–3582.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Lotto, A. J., and Holt, L. L. (**2006**). "Putting phonetic context effects into context: A commentary on Fowler (2006)," Percept. Psychophys. **68**, 178–183.

Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**, 602–619.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)," J. Acoust. Soc. Am. **102**, 1134–1140.

Lotto, A. J., Sullivan, S., and Holt, L. L. (**2003**). "Central locus for non-speech context effects on phonetic identification," J. Acoust. Soc. Am. **113**, 53–56.

Malone, B. J., and Semple, M. N. (**2001**). "Effects of auditory stimulus context on the representation of frequency in the gerbil inferior colliculus," J. Neurophysiol. **86**, 1113–1130.

Malone, B. J., Scott, B. H., and Semple, M. N. (**2002**). "Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques," J. Neurosci. **22**, 4625–4638.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**, 407–412.

Maye, J., Werker, J., and Gerken, L. (**2002**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**, B101–B111.

Miller, L., Escabi, M., Read, H., and Schreiner, C. (**2002**). "Spectral-temporal receptive fields in the lemniscal auditory thalamus and cortex," J. Neurophysiol. **87**, 151–160.

Mitterer, H., Csépe, V., and Blomert, L. (**2006**). "The role of perceptual integration in the recognition of assimilated word forms," Q. J. Exp. Psychol. **59**, 1305–1334.

Movshon, J., and Lennie, P. (**1979**). "Pattern-selective adaptation in visual cortical neurons," Nature (London) **278**, 850–852.

Müller, J. R., Metha, A. B., Krauskopf, J., and Lennie, P. (**1999**). "Rapid adaptation in visual cortical to the structure of images," Science **285**, 1405–1408.

Näätänen, R., Gaillard, A. W., and Mantysalo, S. (**1978**). "Early selective-attention effect on evoked potential reinterpreted," Acta Psychol. **42**, 313–329.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (**2001**). "'Primitive intelligence' in the auditory cortex," TINS **24**, 283–289.

Nelken, I., Rotman, Y., and Bar-Yosef, O. (**1999**). "Response of auditory cortex neurons to structural features of natural sounds," Nature (London) **397**, 154–157.

Newman, R. S., and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**, 540–560.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (**1996**). "Word segmentation: The role of distributional cues," J. Mem. Lang. **35**, 606–621.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (**1999**). "Statistical learning of tone sequences by human infants and adults," Cognition **70**, 27–52.

Saul, A., and Cynader, M. (**1989**). "Adaptation in single units in visual cortex: the tuning of aftereffects in the spatial domain," Visual Neurosci. **2**, 593–607.

Stephens, J. D. W., and Holt, L. L. (**2003**). "Preceding phonetic context affects perception of non-speech sounds," J. Acoust. Soc. Am. **114**, 3036–3039.

Summerfield, Q. (**1975**). "Information-processing analyses of perceptual adjustments to source and context variables in speech," The Queen's University of Belfast, Thesis.

Ulanovsky, N., Las, L., and Nelken, I. (**2003**). "Processing of low-probability sounds by cortical neurons," Nat. Neurosci. **6**, 391–398.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple time scales of adaptation in auditory cortex neurons," J. Neurosci. **24**, 10440–10453.

Wade, T., and Holt, L. L. (**2005a**). "Effects of later-occurring non-linguistic sounds on speech categorization," J. Acoust. Soc. Am. **118**, 1701–1710.

Wade, T., and Holt, L. L. (**2005b**). "Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task," J. Acoust. Soc. Am. **118**, 2618–2633.

Wade, T., and Holt, L. L. (**2005c**). "Perceptual effects of preceding non-speech rate on temporal properties of speech categories," Percept. Psycho-

phys. **67**, 939–950.

Watkins, A. J. (**1991**). "Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **90**, 2942–2955.

Watkins, A. J., and Makin, S. J. (**1994**). "Perceptual compensation for speaker differences and for spectral-envelope distortion," J. Acoust. Soc. Am. **96**, 1263–1282.

Watkins, A. J., and Makin, S. J. (**1996**). "Effects of spectral contrast on perceptual compensation for spectral-envelope distortions," J. Acoust. Soc. Am. **99**, 3749–3757.

On the influence of interaural differences on temporal perception of masked noise bursts

Othmar Schimmel
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands

Armin Kohlrausch[a]
Eindhoven University of Technology, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands,
and Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

In this research, the influence of interaural differences on temporal positioning of the perceived onset of a dichotic broadband noise target in a diotic broadband noise masker was explored. Interaural time or level differences, a combination of these differences, and filtering with head-related transfer functions were applied to broadband noise bursts, to establish lateralization in quiet at center and right positions in the lateral plane. While applying these interaural differences, the subjects' ability to accurately align the noise's onset to the meter of a regular series of diotic reference onsets was investigated for various signal-to-noise ratios. At high target sensation levels, mean onset positions were close to physical isochrony and no systematic differences in temporal positioning due to different interaural cues were observed. At low target sensation levels, positioning accuracy decreased with a decreasing signal-to-noise ratio, and dichotic targets were positioned less accurately than diotic targets. This reduced temporal positioning accuracy could not be attributed to differences in detectability of the targets just above the threshold, nor to a reduction of effective target bandwidth of dichotic broadband targets. From these results, the reduced accuracy of temporal perception in binaural unmasking conditions may be seen as a consequence of temporal integration in binaural processing. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2345913]

I. INTRODUCTION

Fast increases in intensity of signal components, i.e., onsets, are grouped by the human hearing system if they have a level above the detection threshold and occur simultaneously in several spectral regions. Such synchronous onsets are likely to emanate from the same physical source, and are combined to form a perceptual stream or auditory object (Bregman, 1990). The detection of these onsets is, therefore, deemed key to the organization of auditory objects from an acoustic scene.

The perceptual moment of occurrence, i.e., the *perceived* onset, of a sound depends on transient behavior of the auditory system (Schütte, 1978; Zwicker and Fastl, 1999, pp. 271–275). The perceived onset is a function of the sound's duration and its temporal envelope, and may be defined as the moment at which a certain percentage of the maximum of the transient is reached. For a sound with an abrupt onset, the shorter the sound's duration, the stronger its perceived onset is associated with its physical onset.

Directional cues also support and enhance perceptual streaming, although the grouping of signal components is not *initiated* by common interaural time differences only (Hill, 1994; Culling and Summerfield, 1995; Darwin and Hukin, 1999). Interaural differences in time and intensity enable the localization of sound sources, and make the human auditory system more robust in detecting and segregating sounds from

and within complex acoustic scenes. Much of this ability is due to a spatial release from masking, commonly referred to as the *binaural masking level difference*, which is computed as the difference in the detection threshold between a specific binaural condition and a monaural or diotic reference condition. Differences in location of concurrent sound sources provide physical parameters to support differentiation between sounds in *cocktail-party* conditions, in which separating spectrally overlapping signal components would otherwise be difficult. In the present study we investigated whether directional cues of a sound have an influence on perception of a sound's onset, and consequentially may influence the organization of auditory objects.

The most extensively researched phenomenon with regard to interaural differences and the perception of onsets is the *precedence effect*, i.e., the observation that localization of a sound in reverberant environments is mainly determined by the interaural cues of the first-arriving direct sound with decreased influence of later-arriving reflections (Wallach *et al.*, 1949; Litovsky *et al.*, 1999). From measuring the just-noticeable differences in interaural time and intensity of Gaussian noise bursts of less than 5 ms in continuous low-pass filtered noise, the precedence effect can be seen as a result of a temporary degradation of interaural sensitivity between approximately 0.5 and 10 ms after the immediate onset of a sound (Zurek, 1980).

The influence of a signal's sensation level on the precedence effect was investigated by Goverts *et al.* (2000). Using a 2.4-kHz low-pass filtered 5-ms noise burst with fixed but

[a]Electronic mail: armin.kohlrausch@philips.com

ambiguous lateralization cues, they changed the sensation level either by modifying the signal level in silence or adding a noise masker. The onset dominance for lateral perception of the stimulus declined at sensation levels below about 30-40 dB and disappeared completely at 3 dB. Thus, the audibility of a signal's onset influences processing of its directional information.

Conversely, an effect of directional information on temporal perception was established (van Noorden, 1975). The dichotic presentation of an alternating tone sequence of 40-ms tones at 100-ms interonset intervals that, in diotic presentation, yielded a single perceptual stream, resulted in two separate perceptual streams in which the temporal coherence of the original tone sequence was lost. This finding indicated a difference between within-stream and across-stream temporal judgments, induced by directional information, that may influence the perception of a sound's onset.

Furthermore, the effects of directional information on temporal perception may be expected in masking conditions. Binaural unmasking of the target at low signal-to-noise ratios depends on the detection of a change in the interaural correlation (Durlach *et al.*, 1986). This processing requires temporal integration, as it is the change of interaural differences over time and the inconsistency of interaural time differences across frequency that allow the detection of decorrelation. The slow temporal response of the binaural hearing system to changing directional cues, the so-called *binaural sluggishness*, was shown to play a role in binaural detection (Kollmeier and Gilkey, 1990). Directional cues that improve detection through a spatial release from masking may, therefore, oppose the binaural detection of a decorrelating sound's onset.

Our aim in the research presented here was to investigate the influence of interaural differences on onset perception. The question was addressed whether the availability of different directional information for two sounds changes the temporal perception of a sound's onset, compared to a situation without any difference in directional information between two sounds.

II. EXPERIMENT 1

This experiment was defined to measure whether specific differences in interaural parameters between an existing and a newly emerging broadband noise influence the temporal perception of the new noise's onset.

A. Method

1. Procedure

To determine a sound's perceived onset, an experimental procedure was adapted from speech perception research, based on the perceptual center location (Marcus, 1976; Pompino-Marschall, 1989). The perceptual center procedure comprises judging the temporal position of a target sound, and/or moving a target sound along the time axis of a stimulus, relative to isochronous reference sounds in the stimulus, in order to find the desired temporal alignment or perceptual center. These perceptual centers are related to the acoustic
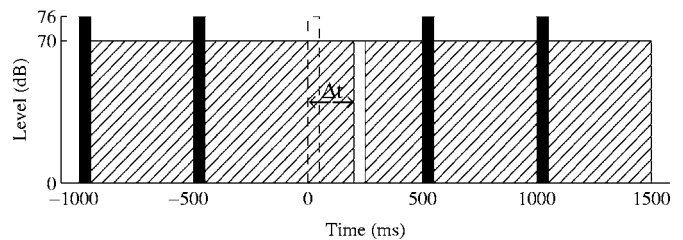
FIG. 1. Example of an acoustic stimulus for the temporal onset positioning experiment. In quiet or against a stationary masker (as depicted here by the cross-hatched area), a randomly positioned target's onset (in white) was to be temporally aligned to the meter of a regular series of reference onsets (in black), to yield the dependent variable adjusted onset position, quantified as the deviation (Δt) from physical isochrony (dashed line at 0 ms).

makeup of a stimulus. In speech sounds, for example, they are determined both by the vowel onset and the duration of preceding consonants (Marcus, 1976).

Because perceptually regular patterns have regular perceptual centers, perceptual isochrony can be achieved with physically isochronous onsets of equal-duration reference and target sounds with identical temporal envelopes. In other words, for perceptually identical (or at least very similar) sounds, the perception of a target sound's onset can be inferred from the accuracy of positioning it to reference sound onsets. A systematic shift from physical isochrony may then be associated to manipulations of the target sound's physical properties. A monotonic relationship between temporal positioning accuracy and the underlying dimension of onset perception is assumed.

The stimulus structure consisted of a regular series of five clearly audible marker pulses of which the third marker pulse was omitted, both in quiet and against a stationary masker. The subjects' task was to temporally align the target's onset, initially positioned randomly around the onset position of the missing third marker pulse, to the meter of the marker pulse onsets (see Fig. 1 for a graphical representation of a stimulus). The time difference between the adjusted temporal position of the target's onset from physical isochrony at the temporal center between onsets of marker pulse two and four was recorded, and served as a quantitative measure for positioning performance. From these measures, temporal positioning bias and accuracy of differently processed targets could be obtained.

The step size for timing adjustments was set to 5 ms, which is roughly the size of the just-noticeable difference in perceptual center changes (cf. Marcus, 1976; Schütte, 1976). The initial random temporal displacements were equally distributed between −125 and 125 ms in steps of 5 ms. The temporal positions the subjects could adjust were limited to the range between −450 and 450 ms around physical isochrony.

Subjects controlled the experiment from a computer keyboard. After an initial playback of the stimulus with a randomly positioned target, the target position could be adjusted by pressing **f** (for 'forward') or **b** (for 'backward') the number of times the target was to be shifted by the step size. Confirmation by pressing **Enter** would start playback of the stimulus with the target at the new temporal position. This procedure could be repeated until subjects were satisfied

with the target onset's fit to the meter of the reference onsets. The final time instant of the target's onset on the time line of the stimulus was then recorded as its perceived moment of occurrence. Feedback was provided after each condition in the form of a graph, representing the followed track from the initial temporal displacement to the final temporal position set. A within-subject experimental design with counterbalanced complete randomization was applied. All conditions were randomly presented twice to each subject, and obtained results were averaged across subjects.

The experiment was conducted in an acoustically isolated listening room at the Philips Research Laboratories in Eindhoven, using a computer running MATLAB software to digitally generate stimuli and automate the experiment and data collection. Digital stimuli were converted to analog signals by a Marantz CDA-94 two-channel 16-bit digital-to-analog converter at a sampling rate of 44.1 kHz, and presented to subjects over Beyer Dynamic DT990 Pro headphones.

2. Stimuli

Three sounds were used: a continuous masker with a level of 70 dB SPL, a number of 50-ms marker pulses with a level of 76 dB SPL, and a 50-ms target with a level of 70 dB SPL. All sounds were broadband noises with independent Gaussian probability distributions, covering the frequency range up to 22.05 kHz, only limited by the frequency response of the headphones. No ramps were applied to the sounds, so they had abrupt onsets and offsets. The masker and the marker pulses were presented diotically, i.e., identically at both ears, to establish lateralization in the center.

The target was presented with different interaural parameters to yield lateralization in quiet at two different positions: one in the center (diotic conditions) and one to the far right (dichotic conditions). This was achieved by the controlled manipulation of interaural time differences, interaural level differences, or by filtering using head-related transfer functions. The following six lateralization conditions were presented: two diotic reference conditions, [C] and [0], and four dichotic conditions, [ITD], [ILD], [ITLD], and [90].

The [C] condition comprised an unprocessed target, identical at both ears. In the [ITD] condition, a time delay was set at 29 samples at 44.1 kHz in the left channel, approximating a 660-μs lag at the left ear. In the [ILD] condition, an interaural level difference was set at 18 dB: target level +9 dB at the right ear, and target level −9 dB at the left ear. The [ITLD] condition combined interaural differences of the [ITD] and [ILD] condition. The diotic [0] and dichotic [90] conditions were computed from head-related impulse responses at 0° and 90° of the HEAD acoustics HMS II artificial head (Blauert *et al.*, 1998).

3. Subjects

Eleven male subjects participated voluntarily in the experiment. All subjects had previous experience in listening tests and reported normal hearing. Training was thought not to be required, because the procedure allowed the subjects to

TABLE I. Mean values and, in brackets, standard deviations of the adjusted temporal onset positions (in ms), for each lateralization condition in quiet and in a noise masker.

	[C]	[ITD]	[ILD]	[ITLD]	[0]	[90]
In quiet	−3.2	−6.2	−9.3	−6.0	−4.3	−5.0
	(7.6)	(12.2)	(9.6)	(11.2)	(11.9)	(11.7)
In noise	2.0	−0.9	8.4	5.0	−0.5	1.1
	(13.0)	(14.5)	(14.6)	(14.1)	(11.8)	(14.1)
Mean	−0.6	−3.5	−0.5	−0.5	−2.4	−1.9

adjust the stimuli for as long as necessary to reach a high confidence level about the temporal onset position before continuing with the next condition.

B. Results

All 22 adjustments (2 adjustments for 11 subjects) for each condition (6 lateralizations at 2 target sensation levels) were pooled, and scanned for severe outliers (onset positions at more than 3 times the interquartile range of the pooled data). Here 2 severe outliers out of the 264 adjustments were removed from the results. Table I shows the mean and standard deviations of the adjusted temporal onset positions. The columns display the six lateralization conditions; the rows indicate the conditions in quiet and in noise.

Table I shows that mean onset positions for all lateralization conditions, both in quiet and in noise, were generally close to physical isochrony and had grand means just below zero. The 95%-confidence intervals of the mean onset positions for each lateralization condition in quiet and in noise revealed that all four dichotic conditions in quiet and the [ILD] condition in noise had means significantly different from physical isochrony.

Because the primary interest was in the temporal positioning *accuracy*, the standard deviations of the adjusted onset positions were used for further analysis. Standard deviations were roughly between 8 and 12 ms in quiet, and between 12 and 15 ms in noise, suggesting an influence of adding the noise masker. A statistical analysis of the difference between variances was performed using paired comparisons by the F ratio, using a 5% level of significance and Bonferroni correction for multiple comparisons. The paired comparisons showed that the introduction of the noise masker significantly reduced temporal positioning accuracy from the condition in quiet only in the [C] conditions ($F_{(21,21)}=2.94$, $p=0.018$). Both in quiet and in noise, results for all six conditions were statistically identical.

C. Discussion

Statistically significant biases from physical isochrony were observed for all four dichotic conditions in quiet. This finding may suggest an effect of lateralization cues. Due to different perceived locations, two independent perceptual streams for markers and target may have been formed in which temporal coherence was compromised, i.e., across-stream judgments of temporal position were worse than within-stream judgments (cf. van Noorden, 1975). However, while these biases were found to be *statistically* relevant,

based on the derived *standard error of the mean*, they were considered to be *perceptually* irrelevant because their differences from physical isochrony were smaller than the *standard deviations* of their underlying distributions. In terms of a d' analysis, all these bias values were below a d' value of one, as was the d' value of the [ILD] condition in noise. Therefore, a systematic bias in temporal positioning for any of the lateralization conditions was not supported by these data. Thus, if the target indeed formed a separate perceptual stream on the basis of different lateralization cues, temporal perception of its onset remained unaffected in the procedure used.

Standard deviations indicated similar temporal positioning accuracy in quiet and in the presence of the noise masker, regardless of interaural differences. An overall comparison of diotic versus dichotic conditions did not reveal a systematic influence of interaural differences on temporal positioning accuracy. In the case of the [C] condition, the unprocessed target was most similar to the diotic markers, so finding the most accurate temporal positioning in quiet here may not be surprising.

With these results in quiet, a baseline for temporal positioning performance was established. The following experiment examined temporal positioning performance at low target sensation levels, where lateralization ability by onset cues was shown to break down (Goverts *et al.*, 2000).

III. EXPERIMENT 2A

This experiment was defined to measure masked threshold levels of diotic and dichotic broadband noise targets in a diotic broadband noise masker, which were needed as reference data for the next temporal positioning experiment at equal target sensation levels.

A. Method

1. Procedure

A three-interval, three-alternative forced-choice procedure with adaptive level adjustment was used in a within-subject experimental design with counterbalanced complete randomization, to measure the individual detection threshold levels for various target conditions. The subjects' task was to identify the interval containing the target. Feedback was provided after each trial. The masker level was 70 dB SPL throughout the experiment, as was the initial target level. The target level was adjusted according to a two-down one-up rule, to track the 70.7%-correct response level (Levitt, 1971). To increase efficiency of the procedure, the initial step size of 8 dB was halved after every second reversal, and was kept constant once it had reached 1 dB. After eight reversals at the 1-dB step size, the median of these eight levels was calculated to estimate threshold. Each condition was repeated four times for each subject. The same equipment and facility were used as in the previous experiment.

2. Stimuli

Two sounds were used: a long-duration masker and a short target. Both target and masker sounds consisted of broadband noise, with independent Gaussian probability dis-
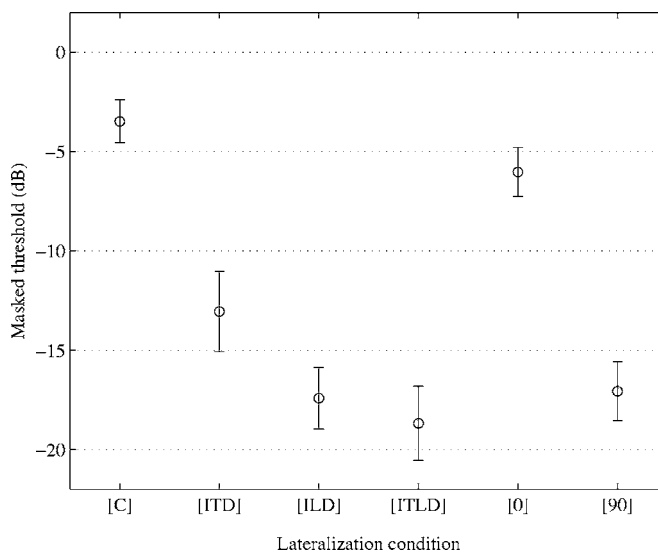
FIG. 2. Mean and standard deviations of the masked threshold levels (in dB, as spectral power ratio between target and masker) for each lateralization condition, based on data of 11 subjects with 4 measurements each. The lower threshold levels of the dichotic conditions ([ITD], [ILD], [ITLD], and [90]) compared to their reference diotic conditions ([C] and [0]) reflect the binaural masking level difference from the application of interaural differences on the target.

tributions, covering the frequency range up to 22.05 kHz, only limited by the frequency response of the headphones. No ramps were applied to the sounds, so they had abrupt onsets and offsets.

The masker had a duration of 400 ms, a level of 70 dB SPL, and was presented diotically. The three masker intervals were separated by 100 ms of silence. The 50-ms noise target was added to the temporal center of one of the maskers. Noise masker waveforms were identical in all intervals within a trial.

The noise target was set at two positions in the lateral plane, one in the center (diotic conditions) and one to the far right (dichotic conditions), by the same manipulations of interaural time differences, interaural level differences, combination of interaural time and level differences, and head-related transfer functions, as described for experiment 1.

3. Subjects

The same 11 subjects participated in this experiment. Again training was thought not to be required, as the procedure allowed straightforward selection of one interval being different from the other two (if detection occurred), without depending on the ability to consistently label the difference.

B. Results

Results of all subjects were pooled, because the between-subject variability (standard deviations across the mean values of the subjects) was similar to the within-subject variability. Figure 2 displays the mean and standard deviation of the masked threshold levels for each lateralization condition, based on data of 11 subjects with 4 measurements each. Threshold levels are expressed in dB as signal-to-overall-noise spectral power ratio between target and masker.

Mean masked threshold levels for the dichotic conditions ([ITD], [ILD], [ITLD], and [90]) were considerably lower than for both diotic reference conditions ([C] and [0]). Compared to the [C] condition, the threshold level for the [ITD] condition was about 9 dB lower, and for the [ILD] condition about 14 dB lower. There were clear differences in the effects of manipulating ITD and ILD; although both cue manipulations caused the target to be lateralized to the far right, thresholds were about 4 dB lower in the [ILD] condition than in the [ITD] condition. In the [ITLD] condition, with both interaural time and level differences present, the threshold was about 1 dB lower than in the [ILD] condition, the condition with the lowest thresholds for the separate lateralization cues. Target presentations with head-related transfer functions yielded a threshold about 3 dB lower in the [0] condition compared to the [C] condition. The average difference in threshold level of about 11 dB between the [0] and [90] conditions was smaller than the difference of about 15 dB between the [C] and the [ITLD] conditions.

C. Discussion

The [C] condition without any interaural time or level differences had the highest detection threshold. Because the signal was identical in both ears, with a bandwidth identical to the masker, detection had to occur on the basis of temporal-profile cues, i.e., an increase in level for the period of the target being present. The calculated level increase of adding a 66.5-dB noise target to a 70-dB noise masker is 1.6 dB. This value is in line with other measured increment-detection thresholds of about 2 dB for 40-ms and 60-ms spectrally matching broadband noises (Heinz and Formby, 1999).

As expected, targets with interaural differences were detected at lower levels. Thresholds for dichotic conditions with only a 660-μs interaural time difference on the target ([ITD]) were 9 dB lower than for the diotic reference condition ([C]). For these stimuli, the cue for detection was likely the interaural decorrelation caused by the addition of the target to the masker (i.e., a change in the perceived width of the masker), rather than a monaural change as in the [C] condition. Binaural interaction for detection of decorrelation needed to be present to account for the 9-dB reduction in detection threshold.

Applying interaural level differences to the target ([ILD]) shifted the detection threshold by about −14 dB. Because of the imposed interaural level differences in the target, the signal-to-noise ratio in the right ear was 9 dB larger than for a diotic target. However, the threshold was 14 dB lower, showing an additional 5-dB shift of the threshold that could not be attributed to the signal-to-noise ratio at the better ear. Again, binaural interaction for the detection of decorrelation needed to be present to account for the additional 5-dB reduction in threshold beyond any better-ear effect. The same was then true for the combination of interaural time and level differences ([ITLD]), which yielded a shift in detection threshold of about 15 dB.

In order to evaluate binaural detection cues, the change in interaural correlation ($\Delta\rho$) for the [ILD] and [ITD] conditions was computed using the formulas from van de Par and Kohlrausch (1995). For the [ILD] condition, this change is by definition constant across the whole frequency range, regardless of the signal bandwidth. For the mean threshold of −17.4 dB, $\Delta\rho$ amounted to 0.05. For the lowest and highest individual threshold values, $\Delta\rho$ amounted to 0.029 and 0.075, respectively. The present authors are unaware of any comparable data on $\Delta\rho$ sensitivity for broadband noise signals of a similar bandwidth as used in the current experiment. The data of Gabriel and Colburn (1981) for detecting a change in correlation for a reference correlation of one showed an increase of $\Delta\rho$ with increasing noise bandwidth. For a noise of 4.5 kHz width, the threshold $\Delta\rho$ values of the two observers amounted to 0.02 and 0.03, about a factor 2 lower than the present results. This remaining factor may be explained by the shorter duration of the signals used in the current experiment (50 ms), compared to the 300-ms signals used in Gabriel and Colburn (1981). Koehnke *et al.* (1986) obtained data for $\Delta\rho$ thresholds of 1/3-octave-wide diotic noise bands centered at 500 and 4000 Hz. The mean values of $\Delta\rho$ were 0.022 and 0.092, respectively. There was considerable intersubject variability in their data, leading to a range from 0.006 to 0.075 for $\Delta\rho$ at 500 Hz, which covers nicely the range found for the broadband experiments in the current data.

The analysis of the [ITD] data in terms of $\Delta\rho$ was more complicated, because for a fixed signal level $\Delta\rho$ depends on the frequency. Therefore, the analysis was performed for critical-band-wide noise at various frequencies. The mean threshold value of −13.0 dB led to the following values of $\Delta\rho$ at 200, 500, and 800 Hz: 0.016, 0.071, and 0.095. The value did further increase toward higher frequencies, but these will be ignored due to the loss of phase locking which will start somewhere around 1 kHz, and the fact that the change in *envelope* correlation, which is the only cue at high frequencies, is much smaller than the change in waveform correlation (van de Par and Kohlrausch, 1995). The values of $\Delta\rho$ in individual low-frequency filters were of similar magnitude as for the [ILD] data, and were close to the values found in the literature. This analysis of the thresholds in terms of $\Delta\rho$ was also supported by the high correlation between individual data for [ITD] and [ILD] conditions, with a correlation coefficient of 0.93. Thus, it can be concluded that detection of the dichotic targets was mediated by changes in the interaural correlation.

In the case of applying head-related transfer functions to the target, i.e., filtering with a *frequency dependent* filter, signal-to-noise ratios were no longer the same in all auditory filters. In the [0] condition, the threshold for detection was about 3 dB lower than in the [C] condition, suggesting a release from masking as an effect of this filtering that introduced considerable amplification and attenuation in different frequency regions of the target sound. For instance, at 1, 3, 5, and above 10 kHz there were peaks in the frequency spectrum of 3 dB or more, relative to the specified level, which may account for the 3-dB reduction in detection threshold.

The detection threshold of the [90] condition was measured to be about −17 dB relative to the 70-dB noise masker, indicating a masking level difference between [0] and [90]

conditions of about 11 dB. Relative level increases of the target in this condition were approximately 8 dB at 1, 7, and roughly between 10 and 16 kHz. A better-ear effect, therefore, could not completely explain this shift in threshold, and again binaural interaction for the detection of decorrelation needed to be present to account for the 11-dB reduction in threshold.

The difference in threshold between the [ITLD] condition and its reference [C] condition was about 15 dB, between the [90] condition and its reference [0] condition about 11 dB. These threshold differences concur with the maximum release from masking (10–15 dB) for out-of-phase tones in diotic broadband noise ($N_0 S_\pi$) at signal frequencies below 500 Hz (cf. Durlach and Colburn, 1978).

IV. EXPERIMENT 2B

This experiment was defined to measure whether specific differences in interaural parameters between an existing and an added broadband noise influence the ability to temporally position the perceived onset of the new noise at low target sensation levels.

A. Method

The same temporal positioning procedure as in experiment 1 was used. The only difference was a systematic variation of the target level.

The same three sounds were used: a continuous diotic masker with a level of 70 dB SPL, a number of 50-ms diotic marker pulses with a level of 76 dB SPL, and a 50-ms target with the same six lateralization conditions as in previous experiments, at various sensation levels. Seven different sensation levels were applied to the noise target: from 0 through 6 dB relative to detection thresholds of the noise target, as measured for each lateralization condition and subject in the previous experiment. By choosing target levels relative to the threshold for each lateralization condition and subject, target sensation levels were equal in all lateralization conditions and data could be compared between conditions and subjects.

The same 11 subjects participated in this experiment.

B. Results

The 22 adjustments (2 adjustments for 11 subjects) for each condition (6 lateralizations at 7 target sensation levels) were pooled, and scanned for severe outliers. Here 11 severe outliers out of the 924 adjustments were removed from the results. Table II shows the mean and standard deviations of the adjusted temporal onset positions. The columns display the six lateralization conditions, the rows display target sensation levels 0 through 6 dB.

Table II shows that onset positions for both diotic conditions were distributed rather symmetrically around physical isochrony, with grand means just above zero, whereas onset positions of all 4 dichotic conditions had negative grand means (i.e., their onsets were positioned earlier than physical isochrony) of about −10 ms. The 95%-confidence intervals of the mean onset positions for each lateralization condition and target sensation level revealed 4 of the 42 con-

TABLE II. Mean values and, in brackets, standard deviations of the adjusted temporal onset positions (in ms), for each lateralization condition and target sensation level.

	[C]	[ITD]	[ILD]	[ITLD]	[0]	[90]
6 dB	−0.7	−5.0	−7.3	−3.4	0.0	−3.4
	(12.7)	(13.2)	(20.1)	(20.2)	(15.8)	(27.9)
5 dB	3.2	7.1	−5.0	−4.5	2.5	−0.2
	(12.5)	(17.6)	(19.3)	(29.0)	(8.4)	(20.8)
4 dB	5.0	5.0	−9.1	−5.2	−3.0	−8.6
	(15.4)	(27.9)	(19.9)	(24.4)	(14.2)	(30.2)
3 dB	−0.7	−5.0	−9.0	−12.5	−2.3	−9.8
	(15.5)	(25.9)	(33.2)	(35.3)	(15.7)	(38.6)
2 dB	1.1	−28.3	3.0	−5.7	1.4	−8.6
	(17.9)	(33.7)	(40.1)	(30.5)	(23.7)	(49.9)
1 dB	−0.7	3.2	−15.0	−19.1	10.9	−22.4
	(26.9)	(53.6)	(65.3)	(50.2)	(57.5)	(67.7)
0 dB	−4.0	−16.1	−29.0	−27.7	9.5	−24.5
	(51.8)	(53.3)	(54.9)	(52.5)	(75.1)	(64.7)
Mean	0.5	−5.6	−10.2	−11.2	2.7	−11.1

ditions to have means significantly different from physical isochrony. Three conditions indicated a bias of about −28 ms ([ITD] at 2 dB, [ILD] at 0 dB, and [ITLD] at 0 dB) and one of about −9 ms ([ILD] at 4 dB).

Because interest was in temporal positioning *accuracy*, the standard deviations were used for further analysis. Figure 3 gives a graphical overview of standard deviations for the six lateralization conditions. The abscissa shows target sensation level (in dB) relative to the detection threshold in the 70-dB noise masker, and the ordinate shows standard deviation (in ms). The six lines connecting different symbols represent results for each lateralization condition. The dashed line at 74 ms represents the calculated standard deviation of the random distribution of initial onset positions (between −125 and 125 ms), i.e., the line to which the standard devia-
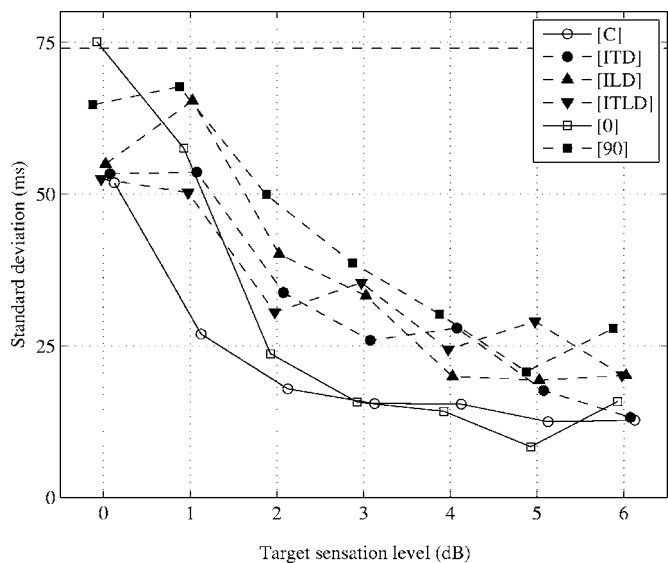
FIG. 3. Standard deviations of the adjusted temporal onset positions (ordinate) versus the target sensation level (abscissa), for six different lateralization conditions. Target sensation levels are relative to the detection threshold in the 70-dB noise masker. The dashed line at 74 ms represents the calculated standard deviation of the random distribution of initial onset positions.

tions would converge if no temporal adjustments were made.

As can be seen on the right side in Fig. 3, standard deviations were lowest at the highest target sensation levels. With a decrease in signal-to-noise ratio between target and noise masker, standard deviations systematically increased for all lateralization conditions. However, standard deviations for dichotic conditions with interaural differences (filled markers) increased faster than those of diotic conditions without interaural differences (open markers). At the detection threshold (0 dB), results had some intrinsic randomness due to the 70.7%-correct level of the procedure followed to establish detection thresholds. Differences in standard deviation between diotic and dichotic conditions were therefore best observed at target sensation levels of 2 through 5 dB. Please keep in mind that results of the dichotic conditions at equal target sensation levels were obtained at much lower *absolute* target levels than in the diotic conditions, reflecting the binaural masking level difference.

A statistical analysis of the difference between variances was performed using the same paired comparisons by the F ratio as in experiment 1. When comparing variances at each target sensation level from 2 through 5 dB, 19 out of 32 comparisons showed significant differences in distributions of adjusted onset positions between the two diotic conditions and the four dichotic conditions ($F_{(19/20/21,20/21)} \geq 3.33$, $p \leq 0.009$). In the same 2 through 5 dB range, the two diotic conditions were statistically identical at each target sensation level, as were the four dichotic conditions.

C. Discussion

Mean differences from physical isochrony, as seen in Table II, were all smaller than their corresponding standard deviations of the underlying distributions, indicating the absence of a perceptually relevant bias in temporal positioning for any of the lateralization conditions. These standard deviations increased toward detection threshold, indicating that temporal positioning accuracy deteriorated when it became increasingly difficult to detect the noise target in the noise masker, regardless of interaural differences. This general deterioration may be due to the transient behavior of the auditory system, in which a lower signal-to-noise ratio results in shallower transients and a consequential disassociation of the perceived onset with the physical onset (cf. Schütte, 1978).

When comparing standard deviations at each target sensation level, significant differences between diotic and dichotic conditions at target sensation levels 2 through 5 dB were found in the paired comparisons of the data shown in Fig. 3. At these target sensation levels, results of the four dichotic conditions (filled symbols) showed worse temporal positioning accuracy than their diotic reference counterparts (open symbols). Results of the two diotic conditions were similar, as were results of the four dichotic conditions. By maintaining equal sensation levels, these findings led to the conclusion that there was an effect of interaural differences on temporal positioning accuracy at low sensation levels.

Three explanations for this influence of interaural differences on temporal positioning accuracy are possible. First, binaural unmasking of the target at low signal-to-noise ratios

TABLE III. Slopes of the psychometric functions fitted to the threshold detection data (in % correct/dB), at the 70.7%-correct level, for each lateralization condition.

	[C]	[ITD]	[ILD]	[ITLD]	[0]	[90]
Slope	16.1	13.6	12.9	12.4	18.4	10.7

depends on detecting interaural decorrelation. The temporal integration required to process the decorrelation of a dichotic target in a binaural unmasking condition, i.e., binaural sluggishness, may then yield more ambiguous and insufficient temporal cues to accurately position the target's onset.

Second, it may be caused by differences in detectability of the noise target at low sensation levels, just above the detection threshold. If there are differences in the slopes of the psychometric functions of the lateralization conditions at the threshold level, this would mean that in lateralization conditions with steeper slopes, (onset) detection at a specific target level just above threshold is more successful, i.e., it occurs at a higher percentage correct of the psychometric function, than in conditions with shallower slopes.

In order to check whether the difference in temporal positioning accuracy between diotic and dichotic listening conditions was attributable to detectability of the target at low sensation levels, psychometric functions were reconstructed for each lateralization condition from adaptive procedure data of the threshold detection experiment. Individual detection data were first transposed in level to have the same 0-dB target level at the 70.7%-correct point, and then pooled across subjects. A Weibull function was fitted to the data, and slopes at the 70.7%-correct level were calculated. Table III shows the slopes of the psychometric functions at the 70.7%-correct level for each lateralization condition.

From earlier research on the relation between the detectability and the energy of a signal for various interaural phase conditions (e.g. Egan *et al.*, 1969), the shapes of the psychometric function were expected to be comparable, except for an attenuation constant reflecting the binaural masking level difference. However, as can be seen in Table III, slopes at the 70.7%-correct level of both diotic conditions were somewhat steeper (16.4 and 18.4%/dB) than those of the dichotic conditions (13.6, 12.9, 12.4, and 10.7%/dB), meaning that in diotic conditions detection at a fixed level just above the detection threshold was slightly easier than in dichotic conditions. These differences between diotic and dichotic condition slopes (at a maximum ratio of about two, between the [0] and [90] conditions) were, nonetheless, too small to conclusively account for differences between diotic and dichotic conditions in temporal positioning accuracy. At target sensation levels of 2 through 5 dB, the percentage correct at these levels was well above 90% in all six lateralization conditions. So, despite the fact that the slopes of the constructed psychometric functions were different, it was concluded that the differences in the temporal positioning accuracy between diotic and dichotic conditions were not simply an issue of detectability.

When comparing slopes of the dichotic conditions with slope data from previous research (cf. Durlach *et al.*, 1986),

values of the broadband signals here (between 10 and 14%/dB) were considerably higher than those of tonal and narrowband signals in noise maskers (between 3 and 8%/dB in the vicinity of the 75%-correct level). It has been argued before (Hicks and Buus, 2000), that psychometric functions for signals covering several auditory filters are steeper than those for signals covering only one auditory filter, like pure-tone signals, and the obtained results seem to support this expectation. In this light, one may also speculate that steeper slopes in the diotic conditions, compared to the dichotic conditions, were a result of a larger number of excited filters in the former conditions, which brings us to the third possible explanation for the difference in positioning accuracy.

Third, the influence of interaural differences on temporal positioning accuracy may be the consequence of a bandwidth effect. It is known that localization based on fine-structure ITDs only operates in the frequency region below approximately 1.5 kHz. For higher frequencies, the phase of the signal at the two ears cannot be compared anymore, and fine-structure information of the input waveform is gradually lost. Furthermore, sensitivity to changes in the interaural correlation decreases above about 1.5 kHz, when correlation changes are no longer detected via the waveform correlation, but only via the envelope correlation (Koehnke *et al.*, 1986; van de Par and Kohlrausch, 1995). It can therefore be assumed that, in the dichotic target conditions applying interaural time differences, high-frequency parts of the broadband target do not contribute much, and low-frequency parts are dominant in detection and lateralization. Because detection in the other dichotic conditions is also mediated by changes in correlation, it may be concluded that only a small part of the broadband lateralized target contributed to detection just above threshold, in contrast to the diotic condition, in which high-frequency channels contribute to detection as well. The hypothesis is that also in temporal positioning at low sensation levels the effective target bandwidth differs between diotic and dichotic conditions, and that temporal positioning becomes more difficult when the target bandwidth is reduced. If this assumption is correct, it is expected that temporal positioning accuracy in a diotic condition at low sensation levels decreases if the target bandwidth is reduced by low-pass filtering. When the target bandwidth in the diotic condition matches the presumed effective bandwidth in a dichotic condition, the accuracy in these conditions should agree. Furthermore, positioning accuracy in the dichotic condition should not be affected by low-pass filtering, because the removed high-frequency parts did not contribute to positioning the broadband target's onset. To investigate a possible target bandwidth effect, the experiments described in the sections before were repeated with low-pass filtered noise targets.

V. EXPERIMENT 3A

This experiment was defined to measure detection thresholds of diotic and dichotic low-pass filtered noise targets in a diotic broadband noise masker, which were needed
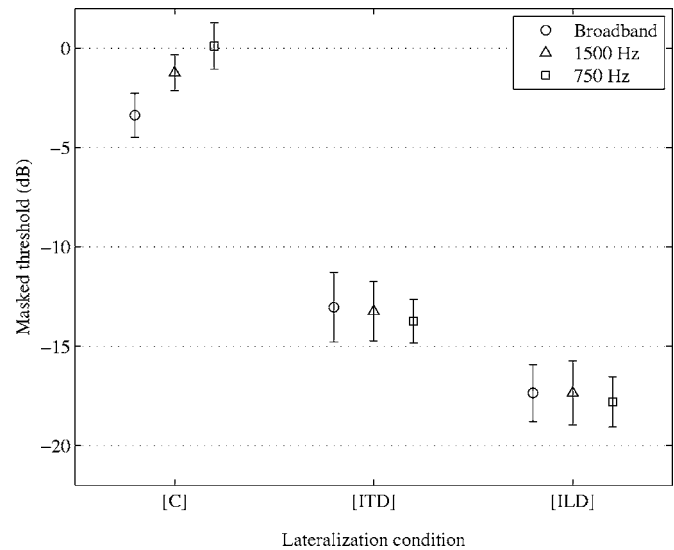
FIG. 4. Mean and standard deviations of the masked threshold levels (in dB, as spectral power ratio between the low-frequency parts of signal and masker) for three lateralization conditions and three target bandwidths, based on data of nine subjects with four measurements each.

as reference data for the next temporal positioning experiment with different target bandwidths at equal sensation levels.

A. Method

The same three-alternative forced-choice procedure with adaptive level adjustment was used to establish individual detection thresholds of low-pass filtered noise targets. Measurements were limited to the [C], [ITD], and [ILD] conditions.

Two different low-pass filtered noise targets were realized by filtering a broadband noise target generated at 70 dB SPL by performing a Fast Fourier Transform, zeroing all signal content above 1.5 kHz or above 750 Hz and applying an inverse Fast Fourier Transform. Target duration was again 50 ms, without on- and offset ramps, but the onset was limited in steepness by the spectral filtering.

This experiment employed a subset of 9 of the 11 subjects from the previous experiments.

B. Results

Results for all subjects were pooled. Figure 4 gives an overview of the mean and standard deviation of the masked threshold levels for each lateralization condition and target bandwidth, based on the data of nine subjects with four measurements each. Threshold levels are expressed in dB as the spectral power ratio between the low-frequency parts of the signal and masker. Displayed are thresholds for the [C], [ITD], and [ILD] conditions for broadband target conditions from the previous threshold detection experiment for the nine subjects involved in the current experiment, and 1500-Hz and 750-Hz low-pass filtered target conditions.

The mean masked thresholds for the low-pass filtered [C] condition were 2.2 dB (1500-Hz low-pass filtered) and 3.5 dB (750-Hz low-pass filtered) higher than for their corresponding broadband condition. Differences for the two di-

chotic conditions were smaller: mean thresholds for the low-pass filtered [ITD] condition were 0.2 dB (1500-Hz low-pass filtered) and 0.7 dB (750-Hz low-pass filtered) lower than for the broadband [ITD] condition, mean thresholds for the low-pass filtered [ILD] conditions were equal to (1500-Hz low-pass filtered) and 0.5 dB lower (750-Hz low-pass filtered) than for the broadband [ILD] condition.

Analysis of variance on the mean thresholds indicated, next to significant differences between lateralization conditions ($F_{(2,72)} = 1437.95$, $p < 0.001$) and between target bandwidths ($F_{(2,72)} = 3.63$, $p = 0.031$), a significant interaction between the lateralization condition and target bandwidth ($F_{(4,72)} = 9.68$, $p < 0.001$). Tukey HSD post-hoc comparisons showed significant differences between all three lateralization conditions, and significant differences between the broadband target condition and the 1500-Hz low-pass filtered target condition. When comparing target bandwidths for each lateralization condition, significant differences were found between the broadband target condition and both low-pass filtered target conditions in the [C] condition.

C. Discussion

Analogous to the previous detection threshold experiment, the noise target in the diotic [C] condition was detected by an increase in the overall level during its presence. The main difference with the previous experiment was that, because of the 1500-Hz and 750-Hz low-pass filtering, also a spectral cue for detection was introduced. Based on earlier experiments about spectral integration (Green, 1958), a theoretical threshold increase of about 1.5 dB for each halving of the number of stimulated auditory filters would be expected when thresholds are expressed as power density. The three target bandwidths used in the current experiment matched about 42 ERB, 19 ERB, and 14 ERB. Thus, the increases observed here agree with the earlier findings cited above.

The detection for the two dichotic conditions was not influenced by the low-pass filtering of the target. This finding suggests that high-frequency parts of the broadband noise target did not contribute to detection in the [ITD] conditions, and is in line with the well-established decrease in BMLD above about 500–700 Hz. Also in the [ILD] condition, low-pass filtering of the noise target did not have a statistically significant effect on the required level for detection. This lack of influence of the high-frequency component may, on a first view, be surprising because interaural level differences were identical in all excited auditory filters. However, detection of the target depended, in addition to level differences, also on changes in the interaural correlation from combining the target and masker. At high frequencies above about 1.5 kHz, only changes in the envelope correlation were available for the binaural analysis. It has been shown by van de Par and Kohlrausch (1995) that for a given target-to-masker ratio, changes in the envelope correlation for Gaussian signals are much smaller than changes in the waveform correlation. Therefore, also in the [ILD] condition, high-frequency channels contributed much less to target detectability than low-frequency channels.

In summary, these detection data support the hypothesis that the "effective" target bandwidth at low sensation levels was indeed much narrower for dichotic targets than for diotic targets.

VI. EXPERIMENT 3B

This experiment was defined to investigate a potential influence of the target bandwidth on temporal onset positioning of broadband noises at low target sensation levels, due to the supposed dominance of the low-frequency region in lateralization and in binaural unmasking.

A. Method

The temporal positioning procedure was repeated for the 1500-Hz and 750-Hz low-pass filtered noise targets, at target sensation levels relative to the new detection thresholds. All experimental settings remained the same as in the previous experiments.

The same noise masker and marker pulses were used: a continuous diotic masker with a level of 70 dB SPL and a number of 50-ms diotic marker pulses with a level of 76 dB SPL, but this time in combination with 50-ms, 1500-Hz, and 750-Hz low-pass filtered targets. The same three lateralization conditions as in the previous experiment were applied to the targets. The noise target was presented at five sensation levels: from 0 through 3 dB relative to detection thresholds of the low-pass filtered noise targets, as measured for each condition and subject, and at 70 dB SPL in quiet.

The same subset of nine subjects participated in the experiment.

B. Results

All 18 adjustments (two adjustments for nine subjects) for each condition (three lateralizations and two bandwidths at five target sensation levels) were pooled, and scanned for severe outliers. Here 8 severe outliers out of the 810 adjustments were removed from the results. Temporal positioning performance of the 1500-Hz and 750-Hz low-pass filtered noise target is shown in Table IV, as are results of the broadband condition from previous experiments for the nine subjects involved in the current experiment. Displayed for each condition are the mean and standard deviation of the adjusted temporal onset positions. The different rows indicate target sensation levels relative to the detection threshold in the noise masker for that particular lateralization condition (0 through 3 dB), and the condition in quiet.

Grand means as shown in Table IV showed that in the diotic conditions both low-pass filtered targets were positioned about 5 ms later than physical isochrony, while in dichotic conditions they were positioned about 5 to 15 ms earlier than physical isochrony. Furthermore, with only one exception, low-pass filtered targets were on average positioned about 5 to 10 ms later than their corresponding broadband targets, in all three lateralization conditions.

The 95%-confidence intervals of the mean onset positions for each lateralization condition, target bandwidth, and target sensation level revealed 8 out of 45 conditions to have means significantly different from physical isochrony. With

TABLE IV. Mean values and, in brackets, standard deviations of the adjusted temporal onset positions (in ms), for each lateralization condition, bandwidth, and sensation level.

	[C]			[ITD]			[ILD]		
	Broadband	1500 Hz	750 Hz	Broadband	1500 Hz	750 Hz	Broadband	1500 Hz	750 Hz
In quiet	−4.2	1.9	2.5	−5.9	−5.6	−4.2	−8.1	−6.4	−5.3
	(7.3)	(13.9)	(11.3)	(13.5)	(19.5)	(19.3)	(7.8)	(21.8)	(17.6)
3 dB	−3.9	5.8	−1.9	−6.5	−22.9	−6.1	−19.4	−12.8	−11.7
	(13.0)	(23.3)	(19.9)	(25.5)	(28.8)	(28.3)	(38.3)	(35.8)	(30.0)
2 dB	0.0	13.1	6.1	−12.2	−3.9	−3.2	1.4	−16.1	−17.5
	(19.4)	(31.3)	(17.8)	(55.2)	(50.2)	(30.4)	(42.3)	(55.2)	(36.3)
1 dB	−2.2	12.2	8.1	−1.4	−5.0	−12.51	−17.2	−3.1	0.3
	(26.5)	(22.5)	(46.3)	(52.5)	(66.9)	(49.4)	(64.8)	(60.3)	(46.1)
0 dB	−10.8	−12.2	20.0	−20.6	−37.8	5.8	−32.5	−13.6	15.0
	(47.0)	(48.9)	(59.5)	(46.5)	(44.6)	(56.6)	(57.5)	(52.2)	(51.9)
Mean	−4.2	4.2	6.9	−9.3	−15.0	−4.0	−15.2	−10.4	−3.8

one exception, however, standard deviations of the mean onset positions of these eight conditions were larger than their difference from physical isochrony. These biases were considered to be perceptually irrelevant, and a systematic bias from physical isochrony for any of the three lateralization conditions was not supported by this data.

As before, standard deviations were taken for further analysis. Figure 5 gives a graphical overview of standard deviations for the three lateralization conditions, for broadband data from previous experiments for subjects involved in the current experiment, and both low-pass filtered target conditions. Low-pass filtered target conditions showed similar patterns of generally increasing standard deviations with a decrease in the signal-to-noise ratio as did the broadband target conditions. For target sensation levels above 0 dB, all diotic conditions had lower standard deviations than in both dichotic conditions, with the exception of the 750-Hz low-pass filtered [ILD] condition at 1 dB.
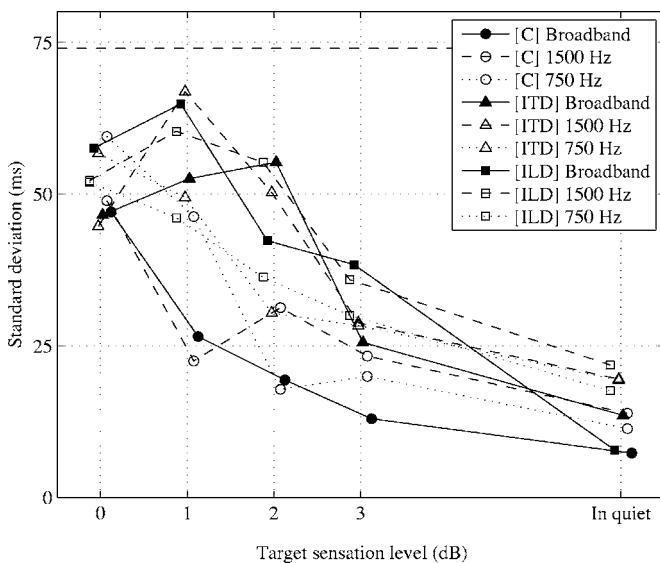
FIG. 5. Standard deviations of the adjusted temporal onset position (ordinate) versus target sensation level (abscissa) for broadband, 1500-Hz and 750-Hz low-pass filtered targets in [C], [ITD], and [ILD] conditions. Target sensation levels are relative to the detection threshold in the 70-dB noise masker. The dashed line at 74 ms represents the calculated standard deviation of the random distribution of initial onset positions.

Paired comparisons by the F ratio were again performed to statistically analyze the difference between variances at sensation levels 1 through 3 dB. When comparing performance between the three target bandwidths in the diotic conditions, only one out of nine comparisons showed a significant difference in temporal positioning accuracy (between both low-pass filtered target conditions at 1 dB, $F_{(15,17)}=4.17$, $p=0.008$). Here 9 out of 12 comparisons between the diotic and their corresponding dichotic low-pass filtered target conditions did not show significant differences. Furthermore, temporal positioning accuracy for all target bandwidths of the dichotic conditions was found to be statistically identical.

C. Discussion

As argued, evidence for a target bandwidth effect should be reflected in two observations: first, a decrease in temporal positioning accuracy at equal low target sensation levels in the diotic condition due to the low-pass filtering, potentially matching the performance of the two broadband dichotic conditions if the effective bandwidth is similar, and second, no effect of low-pass filtering on temporal positioning accuracy in both dichotic conditions.

The first prediction was not supported by the data. Although overall narrowband data in the [C] condition were higher than broadband data, these differences were not significant at the three highest target sensation levels in the masker (1 through 3 dB), where the bandwidth influenced was expected to be seen most clearly. The higher narrowband data did, though, result in the disappearance of significant differences between diotic and dichotic conditions. As for the second prediction, the data showed no significant differences in temporal positioning accuracy between the broadband and both narrowband dichotic conditions, as expected.

From finding no effect of low-pass filtering the target in the diotic conditions, it was concluded that a reduction of the effective target bandwidth could not account for the differences in temporal positioning between diotic and dichotic target conditions at low target sensation levels, which were observed in the broadband condition.

Therefore, the most likely explanation for the observed

differences between diotic and dichotic conditions in temporal onset positioning accuracy at low sensation levels is a systematic influence of binaural processing on perception of a sound's onset. This would mean that in the analysis of a spatial auditory scene, for which a precise temporal analysis of onsets is required (e.g., in speech perception), the advantages in signal detectability from the BMLD are opposed by a reduction in temporal resolution due to the sluggishness of processing binaural detection cues.

VII. GENERAL DISCUSSION

An onset can be detected at the moment a certain level of energy, or change in energy, in one or more frequency parts is reached to make it emerge from its acoustic background, and the new sound can actually be perceived. Though this is certainly true for monaural listening, in binaural listening it is more complex. Current experiments not only showed the expected influence of interaural differences on masked thresholds, but also revealed an influence of interaural differences on the perception of a sound's onset at low sensation levels. Differences in directional cues between an acoustic background and an emerging sound provided detection advantages over having identical directional information, at the cost of a reduction of temporal accuracy in binaural unmasking conditions due to binaural sluggishness. The detection threshold is direction dependent, and so is temporal resolution for perception near this detection threshold.

Onset cues that are known to lead to a strong onset effect, resulting in good localization ability (cf. Zurek, 1993), were already shown to be weakened by a decrease in sensation level and to be ineffective for lateralization at a 3 dB sensation level (Goverts *et al.*, 2000). Here it was argued that temporal perception of dichotic onsets itself was degraded at low sensation levels when compared to the perception of diotic onsets at equal sensation levels. This degradation could not be attributed to differences in detectability just above the threshold. Neither could it be attributed to a reduction of effective target bandwidth of dichotic broadband sounds. The decreasing ability to perceive the onset of a lateralized sound at decreasing signal-to-noise ratios may be seen as a systematic effect of binaural unmasking. At low sensation levels, where unmasking of the target occurs by detecting a change in interaural correlation, the required temporal integration for processing binaural information may yield more ambiguous cues with regard to determining the exact target onset than at high sensation levels.

The difference between diotic and dichotic conditions in perceiving a sound's onset suggests that realization of an auditory object (i.e., the detection of an onset and consequential buildup of a new perceptual stream) in binaural unmasking conditions is influenced by the processing of binaural cues. Although an influence of interaural differences on temporal positioning a perceived onset was established here, the results did not provide further insight in how binaural processing and auditory object formation interact. For instance, localization based on interaural differences can still be independent of, and a parallel process to auditory object formation, as suggested by Woods and Colburn (1992).

VIII. CONCLUSION

Onsets influence binaural processing of interaural differences, and interaural differences influence detection. The aim of the current experiments was to investigate the influence of interaural differences on temporal perception of onsets, a perception that is deemed key to the organization of auditory objects from an acoustic scene. The question was addressed whether the availability of different directional information for two sounds changes perception of an onset, compared to a situation without any difference in directional information between the two sounds.

As expected, binaural cues of interaural time differences, interaural level differences and head-related transfer functions improved the detection of a broadband Gaussian noise target within a continuous broadband Gaussian noise masker. The accuracy of temporally aligning a noise target's onset to the regular meter of reference onsets was influenced by the target's sensation level and directional information. With a decrease in sensation level, the ability to temporally position the target's onset accurately was decreased as a consequence of the transient behavior of the human auditory system. At equal low sensation levels, this temporal positioning accuracy decreased faster in dichotic conditions than in diotic conditions. This reduced temporal accuracy in binaural unmasking conditions is seen as a consequence of binaural processing.

Blauert, J., Brüggen, M., Hartung, K., Bronkhorst, A., Drullmann, R., Reynaud, G., Pellieux, L., Krebber, W., and Sotteck, R. (**1998**). "The AUDIS catalog of human HRTFs," Proc. 16th ICA, pp. 2901–2902.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Durlach, N. I., and Colburn, H. S. (**1978**). "Binaural phenomena," in *Handbook of Perception, Vol. IV Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Chap. 10, pp. 365–466.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1557.

Egan, J. P., Lindner, W. A., and McFadden, D. (**1969**). "Masking-level differences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.

Gabriel, K. J., and Colburn, H. S. (**1981**). "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401.

Goverts, S. T., Houtgast, T., and van Beek, H. H. M. (**2000**). "The precedence effect for lateralization at low sensation levels," Hear. Res. **148**,

88–94.

Green, D. M. (**1958**). "Detection of multiple component signals in noise," J. Acoust. Soc. Am. **30**, 904–911.

Heinz, M. G., and Formby, C. (**1999**). "Detection of time- and bandlimited increments and decrements in a random-level noise," J. Acoust. Soc. Am. **106**, 313–326.

Hicks, M. L., and Buus, S. (**2000**). "Efficient across-frequency integration: Evidence from psychometric functions," J. Acoust. Soc. Am. **107**, 3333–3342.

Hill, N. I. (**1994**). "Binaural processing in a multi-sound environment: The role of auditory grouping cues," Doctoral dissertation, University of Sussex.

Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). "Performance in several binaural-interaction experiments," J. Acoust. Soc. Am. **79**, 1558–1562.

Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Litovsky, R. Y., Colburn, S. H., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Marcus, S. M. (1976). "Perceptual centers," Doctoral dissertation, King's College, Cambridge, UK.

Pompino-Marschall, B. (**1989**). "On the psychoacoustic nature of the P-center phenomenon," J. Phonetics **17**, 175–192.

Schütte, H. (**1976**). "Wahrnehmung von subjektiv gleichmässigem Rhythmus bei Impulsfolgen (Perception of subjectively uniform rhythm in pulse trains)," Fortschritte der Akustik, DAGA '76, 597–600.

Schütte, H. (**1978**). "Ein Funktionsschema für die Wahrnehmung eines gleichmässigen Rhythmus in Schallimpulsfolgen (A functional model of the perception of uniform rhythm in sequences of sound impulses)," Biol. Cybern. **29**, 49–55.

van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Doctoral dissertation, Technische Hogeschool Eindhoven, The Netherlands.

van de Par, S., and Kohlrausch, A. (**1995**). "Analytical expressions for the envelope correlation of certain narrow-band stimuli," J. Acoust. Soc. Am. **98**, 3157–3169.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **62**, 315–336.

Woods, W. S., and Colburn, H. S. (**1992**). "Test of a model of auditory object formation using intensity and interaural time difference discrimination," J. Acoust. Soc. Am. **91**, 2894–2902.

Zurek, P. M. (**1980**). "The precedence effect and its possible role in the avoidance of interaural ambiguities," J. Acoust. Soc. Am. **67**, 952–964.

Zurek, P. M. (**1993**). "A note on onset effects in binaural hearing," J. Acoust. Soc. Am. **93**, 1200–1201.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, 2nd ed. (Springer-Verlag, Berlin).

Temporal and spectral cues in Mandarin tone recognition[a]

Ying-Yee Kong[b]
Hearing and Speech Research Laboratory, Department of Cognitive Sciences,
University of California-Irvine, Irvine, CA 92697

Fan-Gang Zeng
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical
Engineering, Cognitive Sciences, and Otolaryngology-Head and Neck Surgery,
University of California-Irvine, Irvine, CA 92697

This study evaluates the relative contributions of envelope and fine structure cues in both temporal and spectral domains to Mandarin tone recognition in quiet and in noise. Four sets of stimuli were created. Noise-excited vocoder speech was used to evaluate the temporal envelope. Frequency modulation was then added to evaluate the temporal fine structure. Whispered speech was used to evaluate the spectral envelope. Finally, equal-amplitude harmonics were used to evaluate the spectral fine structure. Results showed that normal-hearing listeners achieved nearly perfect tone recognition with either spectral or temporal fine structure in quiet, but only 70%–80% correct with the envelope cues. With the temporal envelope, 32 spectral bands were needed to achieve performance similar to that obtained with the original stimuli, but only four bands were necessary with the additional temporal fine structure. Envelope cues were more susceptible to noise than fine structure cues, with the envelope cues producing significantly lower performance in noise. These findings suggest that tonal pattern recognition is a robust process that can make use of both spectral and temporal cues. Unlike speech recognition, the fine structure is more important than the envelope for tone recognition in both temporal and spectral domains, particularly in noise. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2346009]

I. INTRODUCTION

Acoustic characteristics and their importance for tone recognition have been demonstrated in many tonal languages, including Cantonese (Gandour, 1984), Mandarin (Liang, 1963; Lin, 1988; Whalen and Xu, 1992), and Thai (Abramson, 1978). Results have shown that lexical tones in these languages are distinguished primarily by the change in fundamental frequency (F0) during phonation, although other acoustic properties, such as syllable duration and amplitude contour, also convey tonal information (Whalen and Xu, 1992; Fu *et al.*, 1998; Fu and Zeng, 2000).

Liang (1963) systematically studied the role of spectral cues in Mandarin tone recognition by low- and high-pass filtering the speech stimuli with different cutoff frequencies. He reported that perfect tone recognition can be achieved either with the presence of F0 information with low-pass filtering at 300 Hz or with the harmonic structure with high-pass filtering at 300 Hz. However, in another experimental condition where he only preserved the first formant frequency (300–1200 Hz), performance was still maintained at the 83% level, suggesting the use of formant information in voice pitch perception. Because F0 could still be extracted from harmonics in the first formant frequency range, Liang

also used whispered speech, which contained neither F0 nor harmonic fine structure, to evaluate the contribution of formant information to voice pitch perception. He found that Mandarin tone recognition could be achieved with 60%–80% accuracy (see also Jensen, 1958). Other studies (e.g., Miller, 1961; Howie, 1976), however, showed that tonal contrast was not well preserved with whispered speech. Miller (1961) reported only approximately 40% correct tone recognition in Vietnamese.

In addition to the F0, other information such as temporal envelope was shown to contribute to tone recognition in Mandarin (Fu *et al.*, 1998; Fu and Zeng, 2000; Xu *et al.*, 2002; Whalen and Xu, 1992). Most studies on temporal cues for Mandarin tone recognition concentrated mainly on the envelope cues (e.g., Lin, 1988; Whalen and Xu, 1992; and Fu *et al.*, 1998; Fu and Zeng, 2000; Xu *et al.*, 2002). For Mandarin speech, the amplitude contour cues refer to the slow changes in amplitude and such cues were shown to co-vary with the change of F0 over time in Mandarin. The periodicity cue refers to the fluctuation at a rate, which is directly related to the change in fundamental frequency. Systematically manipulating the amount of the temporal envelope cues, Fu and Zeng (2000) showed that while the duration cues play a minor role in identifying tones, the amplitude contour and periodicity cues contributed significantly to Mandarin tone recognition. Using the noise vocoder processing algorithm described in (Shannon *et al.* 1995), Fu and co-workers (1998) reported that 60%–80% correct tone

recognition in quiet can be achieved, independent of the number of frequency bands. With the amplitude contour cue alone, tone recognition was at 65% correct in the 1-band condition. The additional periodicity cue, however, improved the tone recognition by 10–20 percentage points. The reliance on periodicity cues for tone recognition was also demonstrated by Xu *et al.* (2002), where they reported a trade-off between the amount of spectral information (number of frequency bands) and the low-pass cutoff frequency of temporal envelope extraction. As the low-pass cutoff frequency increased, the number of frequency bands required to achieve the similar tone recognition performance decreased. Xu *et al.* (2002) also created frequency-modulated (FM) pulse trains with F0 contours that mimicked the Mandarin tone contours while eliminating the temporal envelope cue. They reported that without the temporal envelope and duration cues, recognition of vocoder processed FM patterns was poor.

A recent study by Xu and Pfingst (2003) investigated the relative contribution of temporal envelope and fine structure (Hilbert definition) to Mandarin tone recognition. Using the chimearic processing algorithm developed by Smith *et al.* (2002), Xu and co-workers further supported the differential contribution of temporal envelope and fine structure to speech and pitch perception (Smith *et al.*, 2002; Nie *et al.*, 2005). Smith *et al.* found that the temporal envelope cues are sufficient to support speech recognition in quiet, but temporal fine structure is needed for pitch perception. Xu *et al.* (2003) claimed that pitch perception is the common basis for both lexical tone perception and melody recognition by showing that while Mandarin word recognition was determined by the temporal envelope cues, tone recognition was consistent with the temporal fine structure of the chimearized stimuli, with an average of 90.8%, 89.5%, and 84.5% correct for the 4-, 8-, and 16-band conditions, respectively.

The main goal of this study was to systematically investigate the relative contribution of envelope and fine structure cues to voice pitch perception. First, we extended previous studies by examining the envelope and fine structure cues necessary for tone recognition in both temporal and spectral domains. Second, we evaluated the four acoustic cues for Mandarin tone recognition in both quiet and in noise conditions. To our knowledge, no studies have been done on whether reliable Mandarin tone recognition can be achieved in the presence of noise. There are reasons to believe that the effect of noise on tone recognition differs with the type of cue. The terms temporal envelope and fine structure have been defined loosely in the literature. From the speech perception point of view, Rosen (1992) divided temporal envelope cues into three categories, depending on the rate of amplitude fluctuation: (1) envelope (2–50 Hz); (2) periodicity (50–500 Hz), and (3) fine structure (500–10 000 Hz). However, the formal mathematical definition, based on the Hilbert transform (Hilbert, 1912), the temporal fine structure is defined as the instantaneous phase information of the signal. In the present study, the temporal envelope cues were defined as follows: duration and amplitude contours (amplitude fluctuation between 2–50 Hz) and periodicity cues (50–500 Hz). The temporal fine structure was defined as the instantaneous phase information in the signals provided by
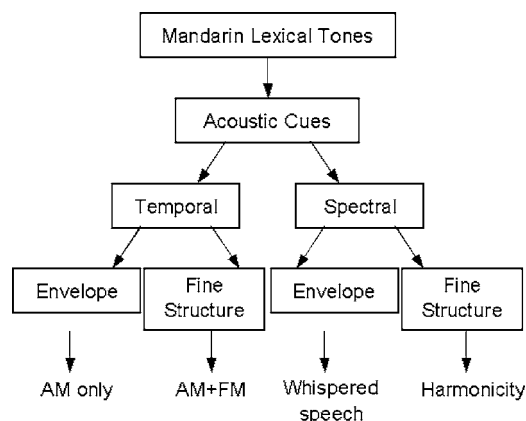
FIG. 1. A block diagram depicts the acoustic cues in the spectral and temporal domains which were evaluated for Mandarin tone recognition.

the Hilbert transform (Hilbert, 1912). As for spectral cues, spectral envelope was defined as the general shape of the spectrum, a smooth curve that passes though the peaks of the spectrum (Hartmann, 1997). For speech stimuli, the spectral envelope represents the filtering properties (or the resonance) in the vocal tract, such as the formant structure of a speech signal. Spectral fine structure, which was also considered as the spectral details, was defined as the detailed frequency components in the power spectrum. The term "spectral fine structure" has been widely used to refer to the spectral details in the power spectrum and we will continue to use this term in this paper. For speech signal, the spectral fine structure represents the source information, which has a harmonic structure for a voiced source and noise structure for an unvoiced source. It should be noted that there are differences between temporal fine structure and spectral fine structure. Consider a simple case of a sinusoidally amplitude-modulated stimulus with a 1000 Hz carrier modulated at a 40 Hz rate; the temporal fine structure is the 1000 Hz carrier of the amplitude-modulated signal, but the power spectrum is composed of three components at 960, 1000, and 1040 Hz.

Figure 1 outlines the four types of acoustic cues that were examined in the study and the stimuli that were used to deliver this information for each experimental condition. We divided the acoustic information into envelope and fine structure in both the temporal and spectral domain. First, for temporal envelope cues, we used the noise vocoder type of processing (Shannon *et al.*, 1995) to manipulate the relative distribution of temporal and spectral information in the speech stimuli. The amount of the temporal envelope information was manipulated by changing the cutoff frequency of the envelope extraction filters, i.e., 50 or 500 Hz in the present study. Previous studies demonstrated that temporal envelope cues contributed significantly to tone recognition in quiet. We hypothesized that the envelope information alone would not be sufficient for tone recognition in noise, due to the nonsaliency of amplitude modulation for pitch perception and also the reduced detectability of amplitude modulation in noise. Second, speech stimuli processed with a novel algorithm developed by Nie *et al.* (2005) were used to assess the importance of temporal fine structure for Mandarin tone recognition. Unlike the noise vocoder processing, which pro-

vided only the amplitude modulation, Nie *et al.* derived additional slowly-varying frequency modulations around the center frequency of each analysis filter and amplitude modulated them with the extracted temporal envelope. We hypothesized that the additional frequency modulation in the speech signal would produce better Mandarin tone recognition. Third, to present only the spectral envelope information in the absence of harmonicity cues, two sets of stimuli were created: naturally recorded whispered speech and synthesized whispered speech using a Linear Predictive Coding (LPC) technique. Lastly, the contribution of spectral fine structure to Mandarin tone recognition was evaluated using stimuli that were the residue of a 14-order LPC processing. While the spectral envelope was presumably flat in this set of stimuli, harmonicity cues were preserved. Based on previous psychophysical findings, we hypothesized that robust voice pitch perception can be determined by the harmonicity cues alone and that spectral envelope cues without the harmonic structure only convey limited voice pitch information.

II. MANDARIN TONE RECOGNITION WITH TEMPORAL ENVELOPE CUES

A. Methods

1. Subjects

Five native Mandarin speakers (S1–S5) participated in the experiment. All of them were females, with ages ranging from 22 to 33. All subjects had normal hearing sensitivity below 20 dB HL at octave frequencies between 125 and 8000 Hz bilaterally.

2. Stimuli

The test material included 100 Mandarin words consisting of 25 syllables with four lexical tones for each syllable (Wei *et al.*, 2004). The original stimuli were recorded from one adult male and one adult female Beijing Mandarin speaker, resulting in a total of 200 words for each test condition. Recording was carried out in a double-walled sound-treated booth and the stimuli were digitally recorded through a preamplifier to the computer at a sampling rate of 22 050 Hz using COOL EDIT PRO (2.0) software. All stimulus waveforms were then adjusted to the same root-mean-squared (rms) amplitude.

Speech-shaped noise was used for the noise conditions. Two samples of speech-shaped noise were created, one for the male speaker and one for the female speaker, by summing the speech tokens (100 words) from either the male or female speaker and then processing them using the 10-order LPC to extract the long-term spectrum of the summed stimuli. While the rms amplitude of speech waveforms was fixed, the rms of the noise waveform varied to achieve the desired signal-to-noise ratios (−10, −5, 0, +5, and +10 dB) tested in this experiment.

The original speech stimuli, combined with the original speech-shaped noise in the noise conditions (male noise combined with the male speech tokens, and female noise with the female speech tokens), were subjected to further processing to contain either temporal envelope cues alone or additional coarse spectral cues via a cochlear implant simu-

lation program similar to that used in Shannon *et al.* (1995). The original broadband (80–8800 Hz) stimuli were first pre-emphasized by a first-order high-pass Butterworth filter at 1200 Hz. The pre-emphasized stimuli were then divided into 1 or 8 frequency bands using sixth-order elliptical bandpass filters. The cutoff frequencies of each band were determined by approximating equal cochlear distance for each band according to the Greenwood map (1990). The corner frequencies for the 1-band processing condition were 80 and 8800 Hz and for the 8-band processing condition were 80, 220, 440, 780, 1300, 2100, 3500, 5500, and 8800 Hz. The temporal envelope was then extracted from each analysis band using two different methods: (1) Hilbert transformation followed by low-pass (LP) filtering (elliptical IIR filters) at 50 or 500 Hz (−6 dB/octave) to smooth out the envelope, or (2) full-wave rectification followed by low-pass filtering (elliptical IIR filters) with cutoff frequency at 500 Hz (−6 dB/octave). While the periodicity cues were preserved in the temporal envelope with the 500 Hz cutoff frequency, it was reduced in the 50 Hz cutoff condition. This envelope signal was then used to modulate white noise, which was then bandpass filtered with the same analysis filters used on the original signal. Stimuli were resynthesized by summing these envelope-modulated narrow-band signals. Note that the temporal envelope fine structure (Rosen's definition) could be preserved in the lower frequency bands with the full-wave rectified envelope but not in the Hilbert envelope.

3. Procedure

All tests were performed in a double-walled sound-treated booth and the stimuli were presented to the listeners monaurally through TDH-49 headphones. The target original and processed speech were presented at a fixed level of approximately 65 dBA. All subjects were first presented with a block of original stimuli followed by four blocks of stimuli processed with the Hilbert envelope (2 spectral conditions ×2 LP cutoffs). Stimuli with higher numbers of frequency bands and higher LP cutoffs were tested first. The most difficult condition (i.e., one-band with 50 Hz LP cutoff) was tested last. Each block of stimuli consisted of one quiet and five different noise conditions. Within each block, the quiet condition was presented first followed by the noise conditions from the highest SNR (10 dB) to the lowest SNR (−10 dB). Therefore, the better performance in the higher SNR conditions compared to the lower SNRs cannot be due to the learning effect. To investigate the effect of the temporal envelope extraction techniques on Mandarin tone recognition, subjects were then further tested with two extra sets of processed stimuli (1-band and 8-band) which contained the envelopes that were extracted using the full-wave rectification and low-pass filtering technique.

The experiment was conducted using a four-alternative, forced-choice procedure, in which four Chinese characters with the same syllable and the tonal marking for each character ("−" for Tone 1, "/" for Tone 2, "∨" for Tone 3, and "\" for Tone 4) were displayed on a computer screen for each trial. The 200 words in the test battery were presented randomly in each experimental condition and subjects were
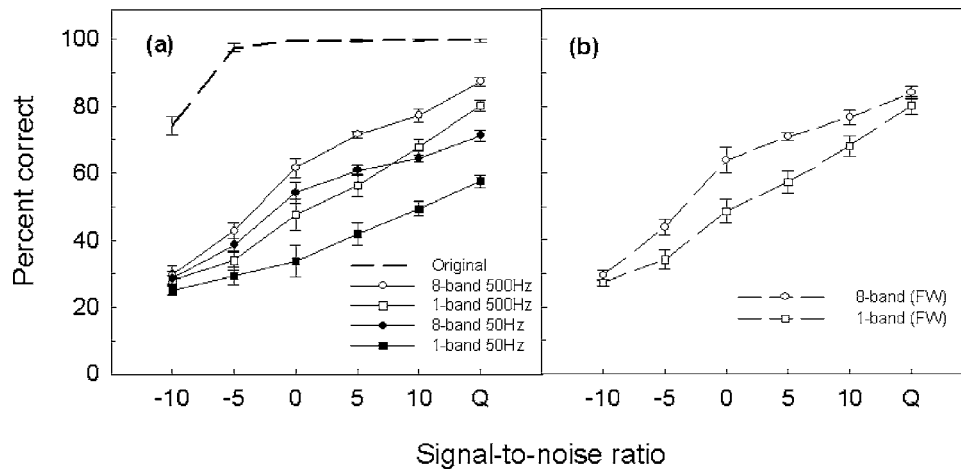
FIG. 2. Average tone recognition scores as a function of signal to noise ratio. Panel (a) shows the tone recognition performance with the Hilbert envelope. Dashed line (- - -) indicates performance with original stimuli. Square symbols represent the 1-band conditions and the circles represent the 8-band conditions. Open symbols indicate the 500-Hz LP conditions and the closed symbols indicate the 50-Hz LP conditions. Panel (b) shows the tone recognition performance with the full-wave rectified envelope in the 8-band (circles) and 1-band (squares) conditions. The vertical bars represent the standard error of the mean.

asked to choose the Chinese characters on the screen that corresponded to the tone they heard. Visual feedback was given after each response.

B. Results

Figure 2(a) shows the average percent correct tone recognition scores as a function of SNR for the original stimuli (dashed line) and for the four processed conditions with the Hilbert envelope (squares: 1-band conditions; circles: 8-band conditions; solid symbols: 50-Hz LP cutoff; open symbols: 500-Hz LP cutoff). On average, tone recognition with original stimuli was maintained at 99%–100% correct from quiet to −5 dB SNR. Performance decreased by 23 percentage points when the SNR was further reduced to −10 dB. Tone recognition was significantly poorer in all of the processed conditions compared to the original stimuli in quiet and at all noise levels ($p < 0.001$).

A three-way repeated-measures ANOVA was performed to examine the following three factors: noise, spectral details (1-vs 8-bands), and temporal details (50- vs 500-Hz LP cutoff). Unlike the original condition in which performance plateaued at −5 dB SNR, tone recognition performance for the four processed conditions increased monotonically as a function of SNR ($p < 0.001$) from nearly chance at −10 dB SNR for all conditions to 86%, 81%, 71%, and 58% correct in quiet for the 8-band 500-Hz LP, 1-band 500-Hz LP, 8-band 50-Hz LP, and 1-band 5-Hz LP conditions, respectively. Significant interactions were observed between the temporal and spectral details [$F(5,20)=15.0, p<0.01$], the temporal details and SNRs [$F(5,20)=13.7, p<0.01$], and spectral details and SNRs [$F(5,20)=8.5, p<0.01$], indicating that the relative importance of temporal and spectral cues for tone recognition differed in quiet and in noise. In general, tone recognition performance improved with 8 frequency bands compared to a single band [50-Hz LP: $F(1,4)=129.1, p<0.001$; 500-Hz LP: $F(1,4)=43.1, p<0.01$]. The presence of periodicity cues also contributed significantly to tone recognition independent of the amount of spectral details [1-band: $F(1,4)=151.8, p<0.001$; 8-band: $F(1,4)=583.1$,

$p<0.001$]. When both the periodicity and spectral cues were available (8-band 500-Hz LP), tone recognition performance was superior to the other three conditions in both quiet and noise (except for the −10 dB SNR), but was the worst when both cues were absent (1-band 50-Hz LP). With either the periodicity (1-band 500-Hz LP) or spectral (8-band 50-Hz) cue, the pattern of performance differed in quiet and in noise. In quiet, the 1-band 500-Hz LP condition produced significantly better tone recognition than the 8-band 50-Hz LP condition [$F(1,4)=47.6, p<0.01$], but was worse in the 0 dB [$F(1,4)=14.4, p<0.05$], and −5 dB [$F(1,4)=8.3, p<0.05$] SNR conditions.

Figure 2(b) depicts the average tone recognition with full-wave rectified envelopes. Compared to the results in Fig. 2(a) with the 500-Hz LP conditions, we can see that tone recognition performance with Hilbert envelopes and full-wave rectified envelopes was essentially identical in both quiet and noise [1-band: $F(1,4)=0.2, p>0.05$; 8-band: $F(1,4)=1.0, p>0.05$], suggesting that listeners were unable to use the envelope fine-structure to perceive voice pitch. In general, the difference between Hilbert and full-wave rectified envelopes was less than 4 percentage points in quiet and at all SNRs for both the 1-band and the 8-band conditions.

III. MANDARIN TONE RECOGNITION WITH SLOWLY-VARYING AMPLITUDE AND FREQUENCY MODULATIONS

A. Methods

1. Subjects

Four additional normal-hearing native Mandarin speakers (S6–S9) participated in this experiment. They were three females and one male, with age ranging from 26 to 41 years. All subjects passed the hearing screening at 20 dB HL at frequencies between 125 and 8000 Hz bilaterally.

2. Stimuli

The recorded Mandarin speech tokens from both the male and female speakers in Experiment 1 were used in this

experiment. The original stimuli were processed in two different ways to extract the slowly-varying temporal envelope (amplitude modulation) alone or the amplitude modulation with additional slowly-varying frequency modulation information.

To produce stimuli with amplitude modulation (AM) cues alone, the signal processing scheme in Experiment 1 was used to extract the Hilbert envelope from 1 to 32 frequency bands in one-octave steps, which produced 1, 2, 4, 8, 16, and 32 subbands. The extracted temporal envelopes were then low-pass filtered at 500 Hz to preserve the periodicity cues in the original speech stimuli.

In addition to the amplitude modulation (AM) information, frequency modulation (FM) was extracted using the algorithm (Frequency-Amplitude-Modulation-Encoding Algorithm) developed by Nie *et al.* (2005) (see also Vongphoe and Zeng, 2005; Stickney *et al.*, 2005 for the description of this algorithm). In this algorithm, the slowly-varying AM was extracted in a similar manner as in the noise vocoder (Shannon *et al.*, 1995) described above. The slowly-varying FM was extracted by first removing each analysis band's center frequency through phase-orthogonal demodulators as used in phase vocoders (Flanagan and Golden, 1966). Two low-pass filters were then used to restrict the FM bandwidth and the FM rate. The bandwidth of the FM was restricted to 500 Hz or the bandwidth of the analysis bands whichever was less, and the FM rate was fixed at 400 Hz. To resynthesize, the slowly-varying FM signal from each subband was used to frequency modulate a sinusoid with a frequency equal to the center frequency of the subband and then the AM signal was used to amplitude modulate the frequency modulated sinusoid. The resynthesized AM and FM signals from each band were summed to recover the original signal. Unlike the fine-structure mathematically defined by the Hilbert transform, in which the obtained instantaneous frequency varies rapidly and over a broad range, the FM encoded using this technique is slowly-varying up to 400 Hz and band-limited to 500-Hz or to the critical bandwidth of the analysis bands.

3. Procedure

Presentation of the stimuli and the experimental setup were identical to those described in Experiment 1. All subjects were presented with a block of original stimuli first, followed by six blocks of AM+FM stimuli, and finally were tested with six blocks of AM-only stimuli. The six blocks within each processed condition (AM-only or AM+FM) corresponded to the six different number of frequency bands conditions from 1 to 32 bands and the blocks were presented in order from highest (32 bands) to lowest number of frequency bands (1 band). Similar to Experiment 1, each block of stimuli consisted of the quiet and different noise conditions and the subjects were presented with the quiet stimuli first followed by the noise conditions from the highest SNR to the lowest SNR. Five noise SNR conditions (−10, −5, 0, +5, +10 dB) were tested for AM-only processed stimuli whereas an additional SNR (−15 dB) was tested for both the original and the AM+FM processed stimuli.

B. Results

Figure 3 shows an average tone recognition performance with the original stimuli [the dashed line in top-left panel of Fig. 3(a)], the AM+FM stimuli (open triangles), and the AM stimuli (filled circles). Similar to the results revealed in Experiment 1, the average tone recognition for the four subjects in this experiment with the original stimuli increased from the chance level performance at −15 dB SNR to 73% correct at −10 dB SNR and then plateaued at 99% correct at −5 dB SNR. Figure 3(a) shows the tone recognition performance as a function of SNR with each panel representing a different number of frequency bands. First note that in the 1-band AM-only condition, tone recognition performance decreased monotonically from 79% correct in quiet to 51% at 0 dB SNR and approached the chance level at −5 dB SNR. While tone recognition was essentially identical between the 1-band and 2-band conditions, tone recognition performance improved significantly as the number of frequency bands further increased in quiet [$F(5,15)=25.6$, $p<0.005$] and at all noise levels [10 dB:$F(5,15)=35.5$, $p<0.005$; 5 dB: $F(5,15)=20.2$, $p<0.05$; 0 dB:$F(5,15)=38.0$, $p<0.005$; −5 dB:$F(5,15)=31.3$, $p<0.005$; −10 dB:$F(5,15)=20.8$, $p<0.05$].This finding is consistent with that reported in Experiment 1. Detailed spectral information, as much as 32 frequency bands, was required to produce tone recognition performance close to the original stimuli. Tone recognition with 32 bands decreased from 99% correct in quiet to 62% at −5 dB SNR and finally to 38% at−10 dB SNR. However, the performance difference between the original and the 32-band AM-only stimuli was significant from −10 to 5 dB SNRs [−10 dB:$F(1,3)=376.4, p<0.001$;−5 dB:$F(1,3)$ $=189.6, p<0.005$; 0 dB:$F(1,3)=177.0, p<0.005$; 5 dB: $F(1,3)=49.0, p<0.01$]. The difference was 5, 15, 38, and 36 percentage points for 5, 0, −5, and −10 dB SNRs, respectively. In contrast, with additional FM, only 4 bands were necessary to achieve performance similarly to the original stimuli for both quiet and noise conditions. Significant performance difference between the 4-band AM+FM and the original stimuli was only 2 percentage points at 0 dB SNR [$F(1,3)=49.0, p<0.01$], 13 percentage points at −5 dB SNR [$F(1,3)=35.4, p=0.01$], and 18 percentage points at −10 dB SNR [$F(1,3)=161.5, p=0.001$].

The difference between the AM+FM and the AM-only conditions was also significant ($p<0.05$) in quiet and at all SNRs independent of the number of frequency bands. For a better visual comparison, the same set of data is replotted in Fig. 3(b) in which average scores are shown as a function of number of frequency bands with each panel representing the different SNR conditions. In general, averaged across different numbers of frequency bands, the differences between AM-only and AM+FM conditions increased from 11 to 36 percentage points as the amount of noise increased from the quiet condition to the −5 dB SNR condition. It is noted that performance with 32 frequency bands in the AM-only condition resembles that observed with two bands in the AM+FM condition, suggesting the contribution of FM cues.
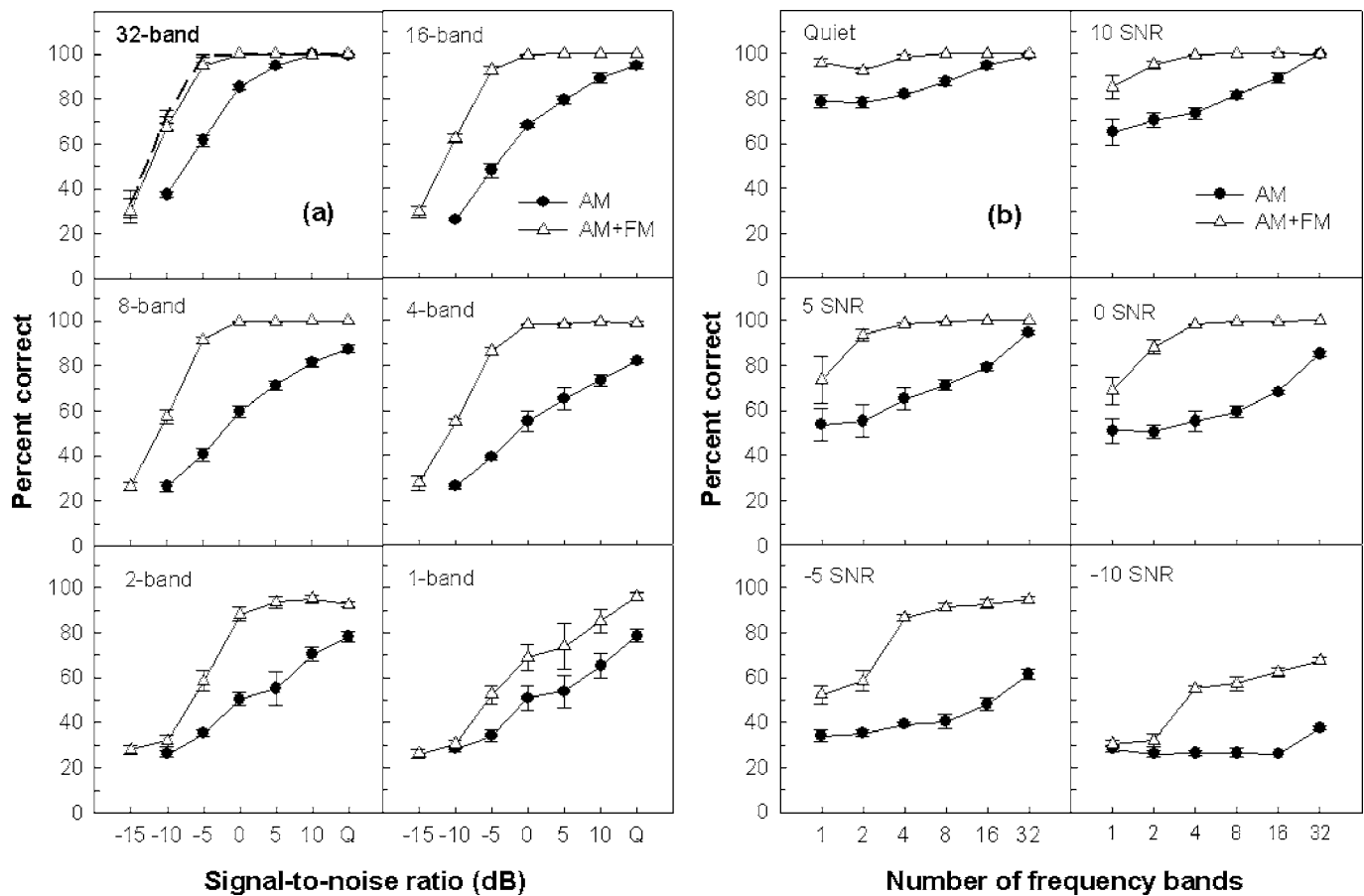
FIG. 3. Average tone recognition with original stimuli [dashed lines in panel (a)], AM alone (solid circles), and AM+FM (open triangles). Panel (a) shows recognition scores as a function of signal-to-noise ratio with individual panels indicating different number of frequency bands. Panel (b) shows the same set of data, but is plotted as a function of number of frequency bands. Individual panels represent different signal-to-noise ratio conditions. The vertical bars represent the standard error of the mean.

IV. MANDARIN TONE RECOGNITION WITH SPECTRAL ENVELOPE CUES

A. Methods

1. Subjects

The same five normal-hearing native Mandarin speakers (S1–S5) that participated in Experiment 1 also participated in this experiment.

2. Stimuli

The same Mandarin word list from Experiment 1 was used in this experiment. Three sets of stimuli were created to evaluate the contribution of spectral envelope and the harmonicity cues to Mandarin tone recognition. Two of the sets, which contained the spectral envelope cues were (1) naturally recorded whispered speech and (2) synthesized whispered speech.

Natural whispered speech was recorded from the same male and female speakers used in Experiment 1, resulting in a total of 200 whispered speech tokens. The recording procedures were identical to those described in Experiment 1. All speech waveforms were adjusted to have equal rms amplitude. Note that the amplitude contour patterns in the whispered speech were similar to those observed in their phonated counterparts. While the harmonic structure of the

phonated speech is replaced by a noise source due to the opening of the vocal folds, the filtering properties (or resonance) of the vocal tract are preserved. In other words, whispered speech contained the formant structures that are similar to its phonated counterparts (von Helmholtz, 1863; Thomas, 1969; Hingashikawa and Minifie, 1999).

Synthesized whispered speech was created by extracting the spectral envelope of the original phonated speech recorded in Experiment 1, using a 14-order LPC technique. Signal processing was performed on every 10-ms time window with a 5-ms overlap. The extracted spectral envelope was then used to modulate with white noise.

To evaluate the contribution of spectral fine structure (i.e., harmonicity cues) to tone recognition, another set of stimuli which contained only the F0 and the harmonics of the original phonated speech was created. The F0 and harmonics of the original phonated speech were extracted as the by-product of the 14-order LPC processing, which was used for creating the synthesized whispered speech. With this processing, the F0, the harmonic, and the phase spectrum of the original stimuli were preserved, but the spectral envelope was flat with no identifiable formant peaks.

3. Procedure

Presentation of the stimuli and the experimental setup were identical to those described in Experiment 1. All sub-
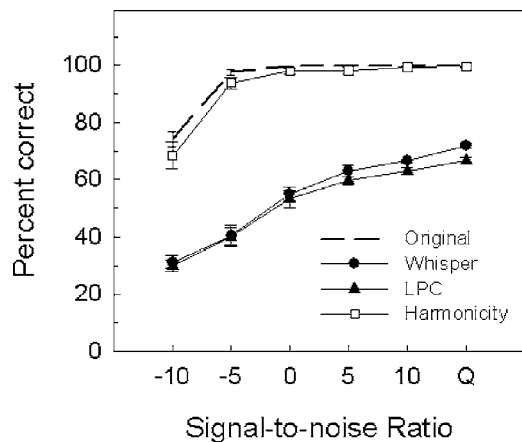
FIG. 4. Mean tone recognition scores with spectral envelope and harmonicity cues. Tone recognition scores with original stimuli (dashed line) from Experiment 1 are replotted in this figure for comparison. Solid circle and triangle symbols represent the natural whispered speech and synthesized whispered speech using the LPC technique, respectively. Open squares represent stimuli with only harmonicity cues. The vertical bars represent the standard error of the mean.

jects were presented with a block of original stimuli first, followed by three blocks of experimental stimuli (natural whispered speech, synthesized whispered speech, and speech with only harmonicity cues). Each block of stimuli consisted of the quiet and five different noise conditions and the subjects were presented with the quiet stimuli first followed by the noise conditions from the highest SNR to the lowest SNR. The noise used for the whispered speech was created by summing all the whispered speech tokens from the male speaker or the female speaker (see noise generating procedures in experiment 1). As for the other two conditions, phonated target speech underwent LPC processing before mixed with noise.

B. Results

Figure 4 shows the average tone recognition scores for the natural whispered speech (circles), synthesized whispered speech (triangles), and speech stimuli with harmonicity cues alone (squares). Mean recognition scores with original stimuli from Experiment 1 are included to facilitate comparison. First note that tone recognition performance with harmonicity cues alone was essentially identical to that with original stimuli at all SNR conditions ($p > 0.05$). Second, tone recognition scores with primarily spectral envelope cues (both natural and synthesized whispered speech) were significantly poorer ($p < 0.0001$) than those obtained with harmonicity cues alone in quiet and at all SNRs. Tone recognition with spectral envelope cues increased monotonically as the SNR increased from chance level performance at −10 dB SNR to 72% and 67% correct in quiet for the natural whispered and synthesized whispered speech, respectively. Third, post-hoc pairwise comparisons with Bonferroni correction revealed that subjects performed similarly ($p > 0.05$) between the natural whispered speech and synthesized whispered speech, except in the quiet condition [$F(1,4)=32.7, p=0.005$] where the natural whispered speech

produced about 5 percentage points better recognition than the synthesized speech.

V. DISCUSSION

A. Comparison with previous studies

Consistent with previous studies (e.g., Fu *et al.*, 1998; Xu *et al.*, 2002), listeners in our study were able to use solely the slowly-varying temporal envelope information, i.e., the amplitude contours and durational cues, to identify the tonal patterns with accuracy of about 60% in the absence of any spectral contents (1-band 50-Hz LP) in quiet. Tone recognition was further improved with additional periodicity cues (1-band 500-Hz LP) by about 20 percentage points. The present results are also in agreement with previous studies of the effect of the number of bands on tone recognition. Both the present and previous results found that tone recognition was essentially the same between one and two frequency bands but improved as the number of frequency bands was increased to 4–8 bands. Using the more challenging stimuli of monosyllabic words with equal durations, Xu *et al.* (2002) found that tone recognition continued to improve up to 12 bands, the highest number of bands tested in their study. In our study, we found further improvement up to 32 bands, with such improvement being particularly apparent in noise.

Xu *et al.* (2002) also reported the existence of a trade-off between temporal and spectral cues for Mandarin tone recognition in quiet. They showed that subjects can achieve a similar tone recognition performance using either a small number of frequency bands combined with a higher temporal envelope cutoff frequency, or a larger number of bands, but with a low envelope cutoff frequency. Here, we observed another relationship—a complementary contribution between temporal periodicity cues and spectral cues for tone recognition between quiet and noise conditions. Note that while the 1-band 500-Hz LP condition produced better tone recognition performance than in the 8-band 50-Hz LP condition in quiet, a reverse pattern was observed in noise. This complementary contribution between periodicity cues and spectral cues suggests that periodicity cues are more important than coarse spectral cues for Mandarin tone recognition in quiet, but they are more susceptible to noise compared to spectral cues.

B. Relative contribution of envelope and fine structure cues to tone recognition

Our study demonstrated that pitch perception was dominated by fine structure rather than the envelope cues. Although tone recognition can be achieved by all four cues in quiet, fine structure cues play a more important role than the envelope cues in noise. In quiet, spectral or temporal envelope cues alone produced relatively low (70%–80% correct) tone recognition performance, but almost perfect performance was observed with spectral or temporal fine structure cues. The difference in performance between envelope and fine structure cues is even greater in noise. From −10 to 10 dB SNR, while tone recognition was still maintained at almost perfect performance level with spectral or temporal fine structure cues, it was 30–60 percentage points lower

with spectral or temporal envelope cues, suggesting that envelope cues are less salient and are more susceptible to noise in voice pitch perception. It should be noted that spectral cues are available in the lower numbers of bands conditions in our AM+FM stimuli and they could be used by our normal-hearing subjects to perceive pitch. Nevertheless, the improved tone recognition in noise with higher numbers of bands (i.e., 16- and 32-bands) is likely to be attributed to the fine timing information within as well as across bands.

The importance of fine structure cues for Mandarin tone recognition is consistent with previous psychophysical findings on pitch perception with pure tones and complex tones. Pitch of a harmonic complex can be extracted based on the place cue (Goldstein, 1973; Terhardt, 1974), or the timing cue (Licklider, 1951; Schouten *et al.*, 1962), or both (Oxenham *et al.*, 2004). Salient pitch percepts can be elicited on the basis of fundamental frequency and the lower resolved harmonics (Ritsma, 1967). Thus, it was not surprising that Mandarin tone recognition was much improved with harmonicity cues compared to spectral envelope cues alone.

In contrast to the fine structures, the saliency of the pitch provided by the temporal envelopes is rather weak, as demonstrated in previous findings on melody recognition with sinusoidally amplitude-modulated noise (Burns and Viemeister, 1981) and modulation frequency discrimination for amplitude-modulated stimuli (Formby, 1985; Grant *et al.*, 1998). The susceptibility of temporal envelope cues to noise in tone recognition could be attributed to the low salience of temporal envelope pitch and the reduced detectability of amplitude modulation in noise.

C. Contribution of formant frequencies to tone recognition

1. Processed speech with temporal envelope cues

Although we have established the role of coarse spectral cues for Mandarin tone recognition, the type of spectral information in this processed speech that enabled the enhancement of tone recognition has not yet been determined. The most important type of spectral information is the F0 and the harmonics. In our recorded speech, the F0 variation patterns were consistent with those reported previously. An acoustical analysis of the F0 of the recorded stimuli was conducted using the correlation algorithm of the TFR™ software.[1] For the male speech, the average F0 of Tone 1 was 174 Hz, the F0 of Tone 2 rose from 113 to 193 Hz, Tone 3 initially fell from 114 to 87 Hz, then rose to 147 Hz, and Tone 4 fell from 189 to 85 Hz. The female speech exhibited similar pitch contours, but with a higher F0 for all tones. In the 8-band AM-only processed speech, the F0 differences among the four tones may not be well represented, due to the fact that the low harmonics were not resolved by the low-frequency analysis filters. Therefore, it was unlikely that better tone recognition with 8-band processing was due to the better encoding of the F0 and the harmonics in the original speech stimuli.

Another type of spectral cue present in speech is the formant structures. Unlike the harmonics, formant information of speech could be coarsely represented in the 8-band

processed stimuli and in turn, vowel recognition performance could be achieved with at least 80% accuracy in quiet (Friesen *et al.*, 2001). However, how is this formant information able to convey differences in Mandarin tones for the same syllable? A careful examination of the formant frequencies of the original recorded single-vowel syllables (/ma/, /pʰɔ/, /kʰɛ/, /ɕi/, /du/, /dʐy/) from the male speaker[2] measured by both the TFR™ software and by a speech synthesis and analysis program PRAAT (Boersma and Weenink, 2005) revealed that:

(1) Formant structures are different for different lexical tones of the same syllable. This finding is supported by the recent articulatory and acoustic data reported in Erickson *et al.* (2004). Using an electromagnetic articulatograph, Erickson and co-workers demonstrated that the jaw and tongue position were significantly more retracted for Tone 3 than for Tone 1 for the vowel /a/, resulting in a higher F1 frequency for Tone 3.
(2) The patterns of difference in formants among tones are not the same for all syllables. In our measurement, F1 frequency was higher for Tone 3 than for Tone 1 for the low vowel, but the pattern of difference was inverted for high and mid vowels. The difference in F1 between Tone 1 and Tone 3 was as large as 100 Hz, particularly for the high vowels. The second formant (F2), however, was consistently higher for Tone 1 than Tone 3 in all vowels, and the difference reached 200 Hz for the back vowels.
(3) Even with a single vowel, the formant frequencies changed considerably over time and the patterns of change differed for different tones. The interesting finding was that the direction of formant changes did not covary with the F0.

The above findings suggest that there may be a possible coupling between the source (vocal folds vibration) and the filter (vocal track configuration) in the articulatory system. This is different from the speech production mechanism proposed in the tradition source-filter theory (Fant, 1960). The source-filter model assumed a linear system in which there was disassociation between the source and the filter.

The considerable difference between F1s among the four lexical tones and the changing frequency of the formants over time for each tone could be represented in an 8-band processed condition. Also, one cannot rule out the possibility that the excitation patterns arising from different filters, due to the overlapping of the filters, may provide additional non-pitch cues for tone recognition. Since the F1 differences among the tones were larger for some vowels (e.g., high vowels) than for others (e.g., mid vowels), it would be of interest to see how tone recognition performance with processed stimuli differed with different syllables. Such information, however, was not recorded in our result matrix during our experiment. Further examination of the effects of vowels on tone recognition is needed.

2. Whispered speech

The contribution of formant frequencies to tone recognition was further supported by our findings with whispered

speech stimuli. Previous studies on whispered speech demonstrated that listeners can (1) match the pitch of whispered vowels to some standard frequency using only the formant frequency information [F1: von Helmholtz (1863); F2: Thomas (1969); F1 and F2: Hingashikawa and Minifie (1999)]; (2) identify speaker's sex with approximately 70%–80% correct performance (Lass *et al.*, 1976; Tartter, 1991); and (3) achieve 60%–80% correct tone recognition in other languages, including Norwegian, Swedish, Slovenian, and Mandarin Chinese (Jensen, 1958). In particular, our results are consistent with Jensen's findings, in which listeners can achieve about 60%–70% correct for whispered speech in quiet. Although most of the studies on whispered speech suggested the contribution of formant information to voice pitch perception, a study by Remez *et al.* (1997) on sinewave speech did not show this relationship. Remez *et al.* (1997) used three sinusoids tracking the lowest three formants of a natural speech and found that normal-hearing listeners could correctly identify the both the speech and speaker. However, they found that the formant (F1 and F2) differences between speakers and the error patterns for speaker identification were poorly correlated, suggesting that listeners did not rely on formant information to identify speakers. They concluded that listeners might be able to use other linguistic features, such as idiolect, dialect, or style differences to identify the talkers. The disagreement between sinewave speech and the whispered speech remains determined.

Acoustical analyses on whispered Mandarin speech were conducted and results showed that the differences of F1 and F2 among the four tones were similar between the whispered speech and their phonated counterparts[3]. These findings were also evidenced in a recent study by Li and Xu (2005). Similar to the phonated speech, both the durational and amplitude contour cues are also available in whispered speech. One might claim that the significant Mandarin tone recognition performance with whispered speech could be attributed to solely the contribution of the temporal envelope cues. However, we argued that the durational and amplitude contour cues alone could not account for relatively high levels of tone recognition in both quiet and noise. The 1-band 50-Hz LP condition yielded a performance level at about 58% in quiet, 14 percentage points lower than the whispered speech and the difference between whispered speech and 1-band 50-Hz LP speech was as large as 21 percentage points at 5 and 0 dB SNR (see Fig. 5). The differences in F1 and F2 among the four tones could be better detected by normal-hearing listeners in the whispered speech compared to the 8-band processed speech. Surprisingly, tone recognition performance was essentially the same between the 8-band 50-Hz LP condition and the whispered speech in both quiet and in noise (see Fig. 5). In addition to the percent correct scores, the error patterns between the 8-band processed and natural whispered speech were similar. These findings suggest that (1) listeners used the same cues to perform the task in both conditions, and (2) well-defined formant information, as in whispered speech, was not necessary for Mandarin tone recognition.

The further improvement beyond 8 frequency bands, i.e., the 16- and 32-band conditions in Experiment 2, may be
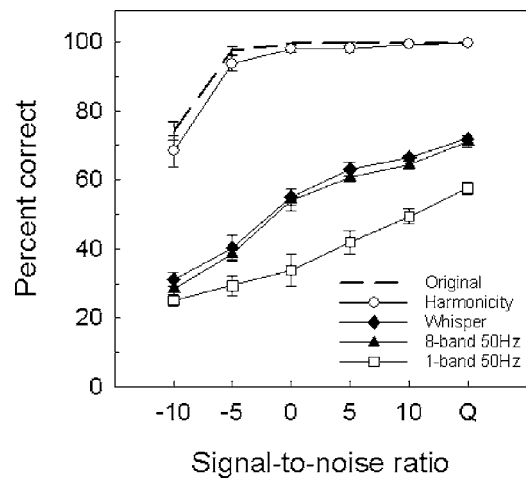
FIG. 5. Mean tone recognition scores for original (dashed line), natural whispered speech (solid diamonds), processed stimuli with 8-band 50-Hz LP (solid triangles), 1-band 50-Hz LP (open squares), and stimuli with harmonicity cues alone (open circles). The vertical bars represent the standard error of the mean.

attributed to the resolvability of the harmonics. Qin and Oxenham (2005) investigated the F0 discrimination limen (F0DL) for a harmonic complex processed by a noise vocoder similarly to Shannon *et al.* (1995). The F0DLs were found similar from 1 to 8 frequency bands and improved considerably in the 24- and 40-band conditions. They concluded that the better F0DLs with the higher number of frequency bands were associated with the spectral cues, and the poorer F0DLs with fewer number of frequency bands were the result of non-salient pitch associated with the envelope cues.

D. Implication for cochlear-implant design

Except for the analog processing strategy, current cochlear implant processing provides only the temporal envelope information from several frequency bands. Both temporal and spectral fine structures are not explicitly encoded in these processing schemes. A study by Wei *et al.* (2004) on Mandarin tone recognition in cochlear-implant listeners measured performance as a function of number of electrodes. They revealed that implant listeners performed only at average 57% correct with a range of 25%–71% correct when listening with a 20-electrode map, considerably lower than the normal-hearing performance even at the 1-band AM-only conditions observed in the present study. Although, the appropriateness of using standard noise-band vocoders as an acoustic model for pitch perception in cochlear implants is being questioned (Laneau *et al.*, 2006), we would still expect implant listeners to be able to perform better than the 1-band AM-only condition given the potential limitation on envelope-periodicity pitch cues in the vocoder. The inability of cochlear-implant listeners to perceive lexical tones was also documented in other languages. Ciocca *et al.* (2002) studied a group of early-deafened Cantonese-speaking cochlear-implant children. They reported that very few of their cochlear implantees performed above the chance level in a two-alternative choice tone identification task.

Improving current cochlear-implant performance in pitch perception can be accomplished by better encoding the temporal and/or spectral fine structures. The present study on Mandarin tone recognition with additional slowly-varying frequency modulation showed significant improvement in Mandarin tone recognition even with a fewer number of frequency bands and also in noisy conditions. One potential problem of adding temporal fine-structure in cochlear implants is that the timing information may not be mapped at the right place along the cochlea in electric hearing, which in turn would still not elicit the salient pitch necessary for voice pitch and music perception. Oxenham *et al.* (2004), using transposed stimuli that resembled the low-frequency pure tones, showed that frequency discrimination with the transposed tones, presented at a higher-frequency place, were considerably poorer than for the pure tones.

Pitch perception in cochlear implants can also be improved with better frequency-to-electrode mapping and/or increasing the number of functional channels. Our results, together with Qin and Oxenham (2005), showed that tone recognition and F0 discrimination improved as increasing the number of frequency bands.

VI. CONCLUSIONS

Mandarin tone recognition was measured in a group of normal-hearing listeners with the goal to understand the underlying mechanisms for voice pitch perception. Four types of speech stimuli were created to evaluate the relative contributions of temporal and spectral envelope and fine structure cues to tone recognition. The main findings are:

(1) Listeners can achieve nearly perfect tone recognition performance with spectral or temporal fine structure cues alone in quiet, but only 70%–80% correct with the envelope cues alone. In noise, the performance difference between fine structure and envelope cues can range from 30 to 60 percentage points from −10 to 10 dB SNR.
(2) While envelope cues, either in the spectral or temporal domain, are sufficient for speech recognition in quiet, fine structure is critical for voice pitch perception in noise.
(3) There exists a complementary contribution of temporal and spectral envelope cues to tone recognition. While the 1-band 500-Hz LP condition produced better tone recognition than the 8-band 50-Hz LP condition in quiet, a reverse pattern was observed in noise. This suggests that temporal envelope cues are susceptible to noise, but spectral cues are more resistant to noise.
(4) Differences in formant frequencies were observed among the four tones for the same vowel in Mandarin speech, suggesting a potential coupling between F0 and formant frequencies.

ACKNOWLEDGMENTS

[1]TFR: Time Frequency Representation software Copyright ©1996, 2000 AVAAZ Innovations Inc.
[2]Measurements of formants for the female speaker were somewhat unreliable due to the high F0.
[3]The formants in the syllable /ɕi/ in the whispered speech could not be reliably obtained due to the excessive noise produced by the speaker.

Abramson, A. S. (**1978**). "Static and dynamic acoustic cues in distinctive tone," Lang Speech **21**, 319–325.

Boersma, P., and Weenink, D. (2005). "Praat: doing phonetics by computer (Version 4.3.14)" (Computer program). Retrieved May 26, 2005, from http://www.praat.org/.

Burns, E. M. and Viemeister, N. F. (**1981**). "Played again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (**2002**). "The perception of Cantonese lexical tones by early-deafened cochlear implantee," J. Acoust. Soc. Am. **111**, 2250–2256.

Erickson, D., Iwata, R., Endo, M., and Fujino, A. (2004). "Effect of tone height on jaw and tone articulation in Mandarin Chinese," presented at the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, Hague).

Flanagan, J. L., and Golden, R. M. (**1966**). "Phase vocoder," Bell Syst. Tech. J. **45**, 1493–1509.

Formby, C. (**1985**). "Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners," J. Acoust. Soc. Am. **78**, 70–77.

Friesen, L. M., Shannon, R. V., Basken, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J. and Zeng, F.-G. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Fu, Q.-J. and Zeng, F.-G. (**2000**). "Identification of temporal envelope cues in Chinese tone recognition," Asia Pacific Journal of Speech, Language, and Hearing **5**, 45–57.

Gandour, J. (**1984**). "Tone dissimilarity judgments by Chinese listeners," J. Chin. Linguist. **12**, 235–260.

Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Grant, K. W., Summers, V., and Leek, M. R. (**1998**). "Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 1051–1060.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hartmann, W. M. (**1997**). *Signals, Sound and Sensation (Modern Acoustics and Signal Processing Series)* (Springer-Verlag, New York).

Higashikawa, M. and Minifie, F. D. (**1999**). "Acoustical-perceptual correlates of 'whisper pitch' in synthetically generated vowels," J. Speech Lang. Hear. Res. **42**, 583–591.

Hilbert, D. (**1912**). *Grundzuge einer Allgerneinen Theorie der linearen Integralgleichungen (Foundations of the General Theory of Linear Integral Calculus)* (Teubner, Leipzig).

Howie, J. M. (**1976**). *Acoustical Studies of Mandarin Vowels and Tone* (Cambridge University Press, Cambridge, England).

Jensen, M. K. (**1958**). "Recognition of word tones in whispered speech," Word **14**, 187–196.

Laneau, J., Moonen, M., Wouters, J. (**2006**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (**1976**). "Speaker sex identification from voiced, whispered, and filtered isolated vowels," J. Acoust. Soc. Am. **59**, 675–678.

Li, X-L., and Xu, B.-L. (**2005**). "Formant comparison between whispered and voiced vowels in Mandarin," Acta. Acust. Acust. **91**, 1079–1085.

Liang, Z.-A. (**1963**). "Hanyu Putonghua zhong shengdiao de tingjiao bian-ren yiju (The auditory basis of tone recognition in Standard Chinese)," Acta Phys. Sin. **26**, 85–91.

Licklider, J. C. (1951). "A duplex theory of pitch perception," Experientia **VII**, 128–134.

Lin, M.-C. (1988). "Putonghua shengdiao de shengxue texing he zhijiao zhengzao. (The acoustic characteristics and perceptual cues of tones in Standard Chinese)," Zhongguo Yuwen (Chinese Language) **204**, 182–193.

Miller, J. D. (**1961**). "Word tone recognition in Vietnamese whispered speech," Word **17**, 11–15.

Nie, K., Stickney, G., and Zeng, F.-G. (**2005**). "Encoding frequency modu-lation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Qin, M. K. and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder pro-cessing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.

Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). "Talker identifica-tion based on phonetic information," J. Exp. Psychol. Hum. Percept. Per-form. **23**, 651–666.

Ritsma, R. J. (**1967**). "Frequency dominant in the perception of the pitch of complex sounds," J. Acoust. Soc. Am. **42**, 191–198.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Stickney, G. S., Nie, K., and Zeng, F. G. (**2005**). "Contribution of frequency modulation to speech recognition in noise," J. Acoust. Soc. Am. **118**, 2412–2420.

Tartter, V. C. (**1991**). "Identifiability of vowels and speakers from whispered syllables," Percept. Psychophys. **49**, 365–372.

Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.

Thomas, I. B. (**1969**). "Perceived pitch of whispered vowels," J. Acoust. Soc. Am. **46**, 468–470.

von Helmholtz, H. L. J. (**1863**). *On the Sensation of Tone as a Physiological Basis for the Theory of Music (translated by Alexander Ellis)* (Dover, New York **1954**).

Vongphoe, M. and Zeng, F.-G. (**2005**). "Speaker recognition with temporal cues in acoustic and electric hearing," J. Acoust. Soc. Am. **118**, 1055–1061.

Wei, C.-G., Cao, K., and Zeng, F.-G. (**2004**). "Mandarin tone recognition in cochlear-implant listeners," Hear. Res. **197**, 87–95.

Whalen, D. H. and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.

Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.

Xu, L. and Pfingst, B. E. (**2003**). "Relative importance of temporal envelope and fine structure in lexical-tone perception," J. Acoust. Soc. Am. **114**, 3024–3027.

Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds

Zhaoyan Zhang, Juergen Neubauer, and David A. Berry

UCLA School of Medicine, 31-24 Rehabilitation Center, 1000 Veteran Avenue, Los Angeles, California 90095-1794

In a single-layered, isotropic, physical model of the vocal folds, distinct phonation types were identified based on the medial surface dynamics of the vocal fold. For acoustically driven phonation, a single, in-phase, x-10 like eigenmode captured the essential dynamics, and coupled with one of the acoustic resonances of the subglottal tract. Thus, the fundamental frequency appeared to be determined primarily by a subglottal acoustic resonance. In contrast, aerodynamically driven phonation did not naturally appear in the single-layered model, but was facilitated by the introduction of a vertical constraint. For this phonation type, fundamental frequency was relatively independent of the acoustic resonances, and two eigenmodes were required to capture the essential dynamics of the vocal fold, including an out-of-phase x-11 like eigenmode and an in-phase x-10 like eigenmode, as described in earlier theoretical work. The two eigenmodes entrained to the same frequency, and were decoupled from subglottal acoustic resonances. With this independence from the acoustic resonances, vocal fold dynamics appeared to be determined primarily by near-field, fluid-structure interactions. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354025]

I. INTRODUCTION

A primary prerequisite for self-sustained oscillation of the vocal folds is that the net transfer of energy from the airflow to the tissue be sufficient to overcome frictional forces (Ishizaka and Matsudaira, 1972; Stevens, 1977; Broad, 1979; Titze, 1988). In order for this transfer of energy to occur, the tissue velocity must be approximately in-phase with the driving forces. Titze (1988) argued that this condition may be facilitated by (1) an inertive vocal tract, (2) the superposition of two or more lower-order eigenmodes which propagate a mucosal wave up the medial surface of the vocal fold, or (3) some combination of the previous two conditions.

For a one-mass model, or any model which does *not* propagate a mucosal wave, it has been shown that self-sustained oscillations can only be achieved when the vocal folds entrain with one of the acoustic resonances of the *vocal tract*, a condition which is facilitated by an *inertive supraglottal tract* (Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972; Titze, 1988). Alternately, it has been shown that self-sustained oscillations of a one-mass model may also be achieved when the vocal folds entrain with one of the acoustic resonances of the *subglottal tract*, a condition which is facilitated by a *compliant subglottal tract* (Fletcher, 1993; Zhang *et al.*, 2006). Because of the strong dependence on the acoustic resonances of the vocal system, we refer to such vibrations as *acoustically driven modes of phonation.* As a rule, such modes of phonation are *undesirable* for voice and speech because they tend to be plagued by frequency jumps and involuntary voice breaks, as the vocal folds entrain to distinct acoustic resonances of sub- or supraglottal tracts.

In contrast, the two-mass model (Ishizaka and Flanagan, 1972), or any model which propagates a mucosal wave, tends to oscillate independently of the acoustic resonances. Instead, its dynamics are governed primarily by a near-field fluid-structure interaction. We refer to such vibrations as *aerodynamically driven* modes of phonation. In this mode of phonation, two or more eigenmodes of the vocal fold tissues (Fig. 1) entrain or phase lock so as to create favorable aerodynamic conditions for phonation. Usually the strongest eigenmode creates an alternating convergent/divergent glottis near the top of the glottal airway [Fig. 1(c)], which makes it intimately associated with the glottal aerodynamics, e.g., a convergent glottis is associated with a relatively high intraglottal pressure, and a divergent glottis is associated with a relatively low intraglottal pressure (Titze, 1988). In the literature, this eigenmode has been referred to as an *x*-11 mode [which according to the nomenclature of Titze and Strong (1975), Berry *et al.* (1994), Berry and Titze (1996) is in the form of x-$n_y n_z$, where x refers to a medial-lateral mode of vibration, and n_y and n_z represent the number of half-wavelengths in the anterior-posterior and inferior-superior directions, respectively]. The second mode, usually capturing an in-phase medial-lateral motion of a rectangular glottis near the top of the glottal airway [Fig. 1(b)], governs the net lateral tissue velocity in this region, and modulates the glottal airflow. In the literature, this eigenmode has been referred to as an *x*-10 mode. Using the method of empirical eigenfunctions (EEFs), Berry *et al.* (1994) investigated the details of this theory with regard to a finite element model of vocal fold vibration. In that investigation, the tissue dynamics were explained by two dominant eigenmodes. The two eigenmodes exhibited a 1:1 entrainment or phase locking such that the intraglottal pressure (governed by an *x*-11-like mode)
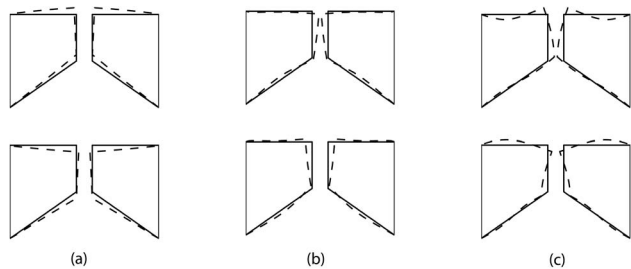
FIG. 1. The three lowest-order eigenmodes of a two-dimensional vocal foldlike structure: (a) eigenmode 1, analogous to the z-10 mode; (b) eigenmode 2, analogous to the x-10 mode, describes in-phase motion along the vertical extent of the medial surface; (c) eigenmode 3, analogous to the x-11 mode, describes out-of-phase motion along the vertical extent of the medial surface. Left and right folds are shown to give an indication of the glottal geometry produced by the corresponding mode. Upper and lower frames indicate extreme positions of eigenmodes, spaced 180° apart in a vibratory cycle. Solid lines indicate equilibrium positions.

FIG. 2. (a) Schematic of the experimental setup, and (b) the hemimodel configuration.

was in phase with the net lateral tissue velocity (governed by an x-10-like mode), thus facilitating energy transfer from the airflow to the tissue.

More recently, Thomson *et al.* (2005) examined these same concepts in the laboratory using a physical model of vocal fold vibration, and a numerical simulation. Based on the results of the numerical simulation (which was driven by an imposed unsteady subglottal pressure as measured from the physical model), they investigated aerodynamic energy transfer mechanisms. In particular, they reported further substantiation of "the hypothesis that a cyclic variation of the orifice profile from a convergent to a divergent shape leads to a temporal asymmetry in the average wall pressure, which is the key factor for the achievement of self-sustained oscillation." However, using the same physical model with a slightly lower Young's modulus, Zhang *et al.* (2006) showed that the physical model usually exhibited self-oscillation *only when entrained to the acoustic resonances of the subglottal tract*, suggesting that the model exhibited *acoustically driven* rather than *aerodynamically driven* modes of phonation. Indeed, Thomson *et al.* (2005) disclosed that the vibration frequency observed in their study was near the half-wavelength resonance of the subglottal tube. Thus, for aerodynamically driven modes of phonation, it would appear that energy transfer mechanisms remain relatively unexplored in the physical model.

Using the physical model of Thomson *et al.* (2005), the purpose of the present study was to (1) attempt to induce aerodynamically driven phonation in the physical model, and (2) to distinguish *aerodynamically driven* and *acoustically driven* modes of phonation based on the medial surface dynamics of the physical model, over a range of subglottal pressures and tracheal lengths. Historically, a variety of studies have suggested that acoustically-driven modes of phonation are analogous to one-mass models of vocal fold vibration (Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972; Titze, 1988; Fletcher, 1993), and that aerodynamically-driven modes of phonation are analogous to any tissue model which propagates a mucosal wave, usually through the superposition of two or more, lower-order eigenmodes (Titze and Strong, 1975; Titze, 1988; Berry *et al.*, 1994; Berry,
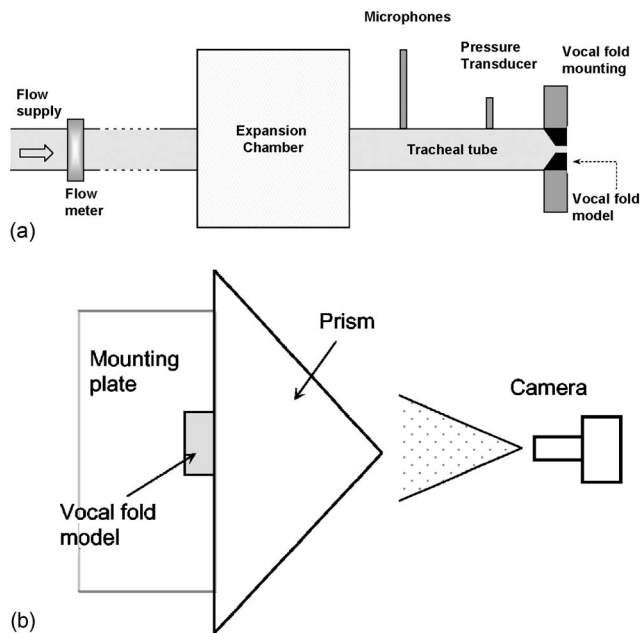
2001; Doellinger *et al.* 2005a, 2005b). Consequently, we hypothesize that aerodynamically and acoustically driven modes of phonation may be distinguished based on the medial surface dynamics of the vocal fold, and by their degree of dependence on the acoustic resonances of the vocal system. In particular, we will seek to substantiate and elucidate this hypothesis with quantitative, experimental data from the physical model.

II. METHOD

A physical replica of the human vocal system [Fig. 2(a)] was constructed using a rubber model of the vocal folds and a uniform (PVC) tracheal tube (2.54-cm inner diameter) connected upstream to an expansion chamber, simulating the subglottal system. The rubber model was a single-layered, cover-only, isotropic, physical model of the vocal folds (Thomson *et al.*, 2005; Zhang *et al.*, 2006). Using a mold, it was created with a two-component liquid polymer solution mixed with a liquid flexibilizer solution. For additional details regarding the fabrication and dynamical properties of the physical model, please see the original papers. Each vocal fold model measured approximately 1 cm in the superior-inferior direction, 1.7 cm in the anterior-posterior direction, and 0.8 cm in the medial-lateral direction [Fig. 3(a)]. The inferior side of each vocal fold had an entrance convergence angle of approximately 60° measured from the inferior-superior axis, yielding an inferior-superior vocal fold thickness of approximately 5.4 mm. Using a 5544 Instron Testing System, the stress-strain relationship of the artificial vocal fold tissue was measured to be nearly linear in the strain range of 0–20 %. Using the stress-strain data, the Young's modulus was calculated to be approximately 11 kPa across this range. The density was 997 kg/m³. The vocal folds were glued into a rectangular groove on the medial surface of two
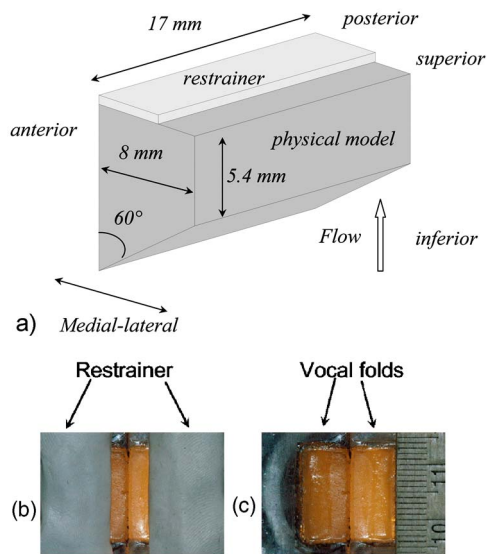
FIG. 3. (a) A sketch of the physical model of the vocal fold, and a superior view of the model both (b) with, and (c) without restrainers.

acrylic plates. The medial surfaces of the two folds were positioned to be in contact so that the glottis was closed when no airflow was applied.

Different lengths of the PVC tracheal tube were used in this study, ranging from 11 to 120 cm. No efforts were made to smooth the transition from the tracheal tube to the physical model, although such an abrupt transition may have generated undesired perturbations in the airflow. The geometry of the transition region was kept unchanged so that its effects remained constant during the course of the experiments. The expansion chamber had an inner cross-sectional area of 23.5×25.4 cm and was 50.8-cm long. The inside of the expansion chamber was lined with a 2.54-cm-thick layer of fiberglass. Cross-sectional area data of the subglottal airway (Ishizaka et al., 1976) have shown that the cross-sectional area function from the trachea and the primary bronchi to the lungs increases abruptly by 10–100 times over a distance of about 6 cm. In this setup, the cross-sectional area from the pseudotrachea section to the expansion chamber increased by a factor of about 117. The expansion chamber was connected to the air flow supply through a 15.2-m-long rubber hose, reducing possible flow noise from the air supply. The acoustic characteristics of the expansion chamber and the flow supply were evaluated using the two-microphone method in a separate study (Zhang et al., 2006) and were found to be similar to an ideal open-ended termination of the tracheal tube for frequencies above approximately 50 Hz. Admittedly, this subglottal system does not exactly reproduce the human subglottal system, particularly the lossy compliance of the lungs. However, the relative simplicity of the system and its acoustical similarity to the human lungs yielded a reliable, controllable model of the subglottal acoustic system, which was essential for the present study.

A hemi-model procedure was used to investigate the medial surface dynamics of the vocal fold model, a technique which has been previously implemented for a variety of other laboratory experiments, including other physical mod-

els of the folds (Titze et al., 1995; Chan et al., 1997), excised larynx experiments (Jiang and Titze, 1993; Berry et al., 2001; Doellinger et al., 2005b), and the in vivo canine laryngeal model (Doellinger et al., 2005a). Figure 2(b) shows a schematic of the hemi-model setup. In this experimental configuration, one of the vocal fold plates was removed and replaced by a glass prism. The prism provided two distinct views of the medial surface of the vocal fold, which was imaged using a high-speed digital camera (Fastcam-Ultima APX, Photron Unlimited, Inc.). A frame rate of 2000 Hz was used with a spatial resolution of 1024×1024 pixels per image. Prior to imaging, graphite powder was sprinkled on the medial surface of the vocal fold to form random dot patterns. In the post-processing stage, such patterns facilitated cross-correlation analysis to compute the medial surface dynamics.

In particular, time-series cross-correlation analysis was performed on the medial surface images using the image-processing package DaVis (LaVision Inc.). A multipass algorithm was used in the cross-correlation analysis. Initial estimates of the displacements were made using a relatively large interrogation window. Next, these estimates were used as input for subsequent analyses which used smaller interrogation window sizes. For each camera view, the correlation analysis yielded the medial surface displacements in image coordinates. Using an appropriate mapping, the three-dimensional, physical coordinates of the displacements were derived from the two sets of image coordinates. The mapping function was determined through means of a calibration process (Hartley and Zisserman, 2000), using a calibration target with known three-dimensional coordinates.

In this case, the calibration target (LaVision Inc., Type no. 2.5) was 25×25 mm with a 2-mm spacing between calibration points in the inferior-superior and anterior-posterior directions, and 0.5-mm spacing between calibration points in the medial-lateral direction. Due to the periodic motion of the medial surface, the correlation analysis was performed with respect to the first frame. Therefore, a Langragian displacement field was obtained as a function of the reference coordinates in the first frame. Ultimately, displacements were computed over a total medial surface area of 16×16 mm, with a grid spacing of 0.22×0.22 mm in the inferior-superior and anterior-posterior directions, respectively. In contrast to previous methods of tracking vocal fold displacements which were semiautomatic (Berry et al., 2001; Doellinger et al., 2005b), the present technique was fully automated.

The subglottal pressure in the tracheal tube was monitored using a probe microphone (B&K 4182), which was mounted flush with the inner wall of the tracheal tube, 5-cm upstream from the vocal fold plates. A pressure tap was also mounted flush with the inner wall of the tracheal tube, 2-cm upstream from the vocal fold plates. The time-averaged transglottal pressure was measured using a pressure transducer (Baratron type 220D). Because no vocal tract was used in this study, the subglottal pressure was equivalent to the transglottal pressure. The volumetric flow rate through the orifice was measured using a precision mass-flow meter (MKS type 558A) at the inlet to the setup. During the experiments, the flow rate was increased from zero to a certain
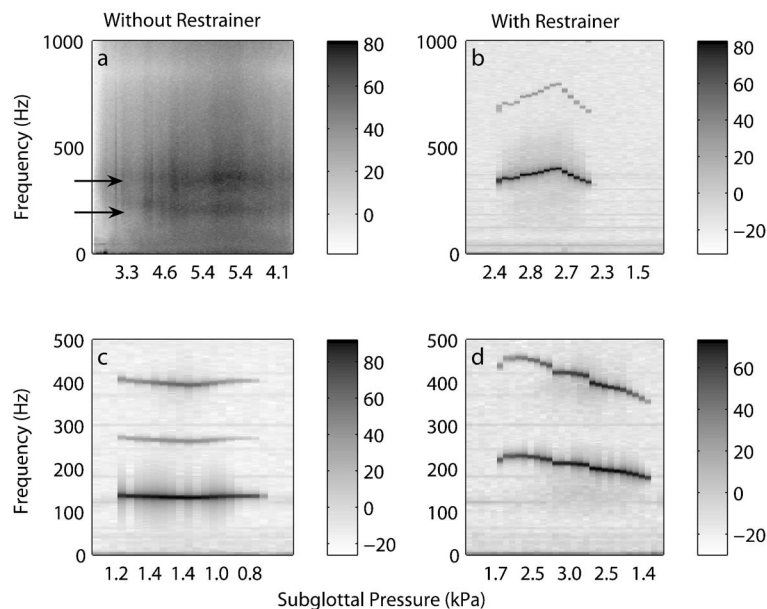
FIG. 4. Sound pressure power spectra as a function of increasing and decreasing subglottal pressure in the hemimodel configuration for: (a) a tracheal tube of length 17.1 cm, without restrainer; (b) a tracheal tube of length 17.1 cm, with restrainer; (c) a tracheal tube of length 60 cm, without restrainer; (d) a tracheal tube of length 60 cm, with restrainer.

maximum value in discrete increments, and then decreased back to zero in discrete decrements (Zhang *et al.*, 2006). At each step, measurement was delayed for an interval of about 4–5 s after the flow rate change, allowing the flow field to stabilize. Sound pressure inside the subglottal tube, flow rate, and subglottal pressure were recorded for a 2-s period. Analog-to-digital conversion of the output signals was performed using a United Electronic Industries Powerdaq board, with 16 bit resolution over a ±10 V measurement range at a sampling rate of 50 kHz.

III. RESULTS

In a companion paper (Zhang *et al.*, 2006), the vibrations of the physical model were studied as a function of tracheal length and subglottal pressure. In that study, the tracheal length was varied from 11 to 325 cm, and subglottal pressure was varied between 0 and 6 kPa. Despite the desire to generate both aerodynamically and acoustically driven modes of vibration, over the range of parameters investigated (0–6 kPa for the subglottal pressure and 11–325 cm for the tracheal length), only acoustically driven modes of vibration were produced by the model, i.e., phonation only occurred when the phonation frequency entrained with one of the subglottal acoustic resonances, and showed little variation with subglottal pressure. Furthermore, as a function of tracheal length, the vocal folds exhibited bifurcations or abrupt jumps to distinct subglottal resonances. Finally, the acoustically driven modes did not occur for typical human tracheal lengths (i.e., approximately 17 cm). Instead, such modes appeared only for tracheal lengths of 30 cm or more. It was also noted that the acoustically driven modes of vibration exhibited significant vertical (inferior-superior) vibrations similar to the eigenmode in Fig. 1(a), a vibration pattern which easily coupled with the subglottal acoustics.

Thus, over the original range of parameters investigated, aerodynamically driven modes of vibration were not observed. In this study, in a further effort to induce aerodynamically driven phonation, several physical constraints

were applied to the vocal fold model to reduce the large vertical vibrations characteristic of the acoustically driven modes of phonation, and thereby reduce source-tract coupling. Indeed, it was hoped that the restriction of the large vertical vibrations would suppress the vertical motion and facilitate the emergence of an *x*-11 eigenmode and mucosal wave propagation, and thereby promote an aerodynamically driven mode of phonation (Titze, 1988). The following constraint produced the desired effect: rubber restrainers were laterally positioned over the superior surface of the vocal fold model [Figs. 3(a) and 3(b)]. The thickness of the restrainers was chosen so that the restrainers were thin enough to avoid disturbance of the near-field flow (approximately 2.5 mm), yet stiff enough to restrain the vertical motion of the vocal fold body. The lateral-medial extent to which the restrainers covered the vocal fold was adjustable. The lateral distance of the uncovered superior portion of the vocal fold roughly approximated the effective lateral depth of vibration of the physical model.

Experiments were conducted on both full and hemi-model configurations. Both configurations produced qualitatively similar vibration patterns. The vibration frequencies between configurations generally differed by a few hertz. Also, phonation threshold pressures were slightly higher (0.1–0.2 kPa) in the hemimodel than in the full model. However, as expected, the threshold airflow in the hemimodel was about half that of the full model. Because of our focus on medial surface dynamics, only results from the hemimodel experiments are discussed below.

A. Phonation frequency

Figures 4(a) and 4(b) compare the power spectra of the subglottal acoustic pressure as a function of the mean subglottal pressure for cases with and without restrainers, for a tracheal tube length of 17.1 cm, which is a typical estimate of the length of the human trachea (Flanagan, 1958; Ishizaka *et al.*, 1976). The subglottal resonance for this tracheal length could be measured directly as dominant energy peaks in the

acoustic power spectra before onset, or estimated from linear acoustic theory using the measured reflection factor of the subglottal system in Sec. II. In this case, the first subglottal resonance occurred around 350 Hz. The power spectra for an unrestrained fold are shown in Fig. 4(a), and the power spectra for a restrained fold are shown in Fig. 4(b). Note that, during the experiments, the mean subglottal pressure was first increased to a maximum value and then decreased back to zero. For the unrestrained fold, phonation did not occur. In this case, two broad horizontal lines can be seen in Fig. 4(a), which correspond to broadband flow noise around the subglottal acoustic resonance of approximately 350 Hz, and the would-be phonation frequency of approximately 210 Hz (small-amplitude vibrations were observed at this latter frequency at higher subglottal pressures than those originally investigated in this study [up to about 7 kPa]). Figure 4(b) shows the restrained case with an effective lateral depth of vibration of 2.2 mm (compared with an effective lateral depth of vibration of 8.0 mm when no restrainer was applied). Phonation onset occurred at a subglottal pressure of approximately 2 kPa. After onset, phonation frequency appeared to vary continuously as a function of subglottal pressure. Although close to the acoustic resonance, the fundamental frequency clearly exhibited variations which were independent of the subglottal resonance.

Figures 4(c) and 4(d) show the acoustic power spectra for unrestrained and restrained folds, respectively, when a longer tracheal tube of approximately 60 cm was utilized. The first subglottal resonance was about 134 Hz. Without the application of the restrainer, phonation was observed. The vocal fold entrained to a subglottal acoustic resonance, and the fundamental frequency exhibited weak dependence on the subglottal pressure [Fig. 4(c)]. Clearly, the physical model vibrated in an acoustically driven mode. However, with the restrainer (with an effective lateral depth of vibration of 5.4 mm), the physical model vibrated independently of the subglottal acoustic resonance (about 134 Hz), and appeared to vary continuously as a function of subglottal pressure in the range of 200–250 Hz, as shown in Fig. 4(d).

Shown in Fig. 5 are the medial surface trajectories of one coronal slice of the medial surface, midway between anterior and posterior extremes of the same physical model with and without restrainers. Without restrainers, large vertical vibrations were observed in the acoustically driven mode of vibration. However, this vertical motion was suppressed when restrainers were applied. Note that lip vibrations in brass instruments exhibit similar vibratory phenomena. For example, the unrestrained vibrations observed in this study roughly correspond to the "swinging door" motion of the vibrating lips, and the vertically restrained vibrations roughly correspond to the "sliding door" motion (Adachi and Sato, 1996; Copley and Strong, 1996).

The fact that the physical model vibrated independently of the subglottal acoustic resonances illustrates that the restrained physical model vibrated in the aerodynamically driven modes of vibration. This suggests that the application of vertical constraints may facilitate the excitation of aerodynamically driven modes of vibration, perhaps by enhancing the near-field fluid-structure energy transfer. The results
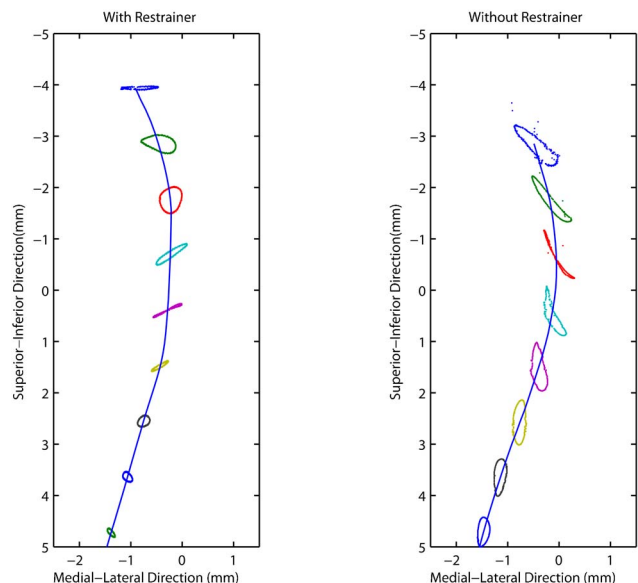
FIG. 5. Anterior view of medial surface trajectories for one coronal slice of the medial surface of the vocal fold, midway between anterior and posterior extremes, for cases with (left) and without (right) restrainers. Because the vibrational amplitudes from the case with restrainers were relatively small, they have been amplified by a factor of three. For a clearer illustration, only trajectories from every fifth grid point are shown along the inferior-superior length. The tracheal tube length was 60 cm.

also suggest that the suppression of vertical motion reduced the coupling of the tissue vibrations with the subglottal acoustics.

B. Spatiotemporal analysis

Figure 6 is a spatiotemporal representation of the medial-lateral component of motion for one coronal slice of the medial surface of the vocal fold model, midway between anterior and posterior extremes. For each subplot, the x axis corresponds to time and the y axis corresponds to the inferior-superior direction. The superior edge of the vocal
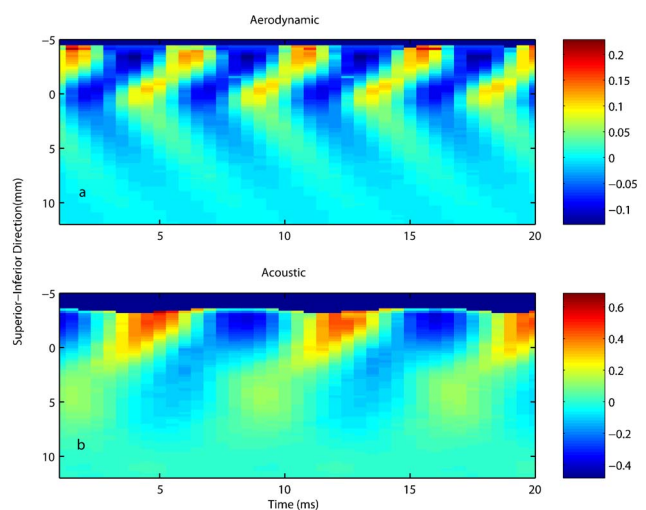
FIG. 6. (Color online) Spatiotemporal plot of the medial-lateral component of the medial surface displacements for one coronal slice of the physical model, for both aerodynamically driven (top) and acoustically driven (bottom) modes of phonation. The tracheal tube length was 60 cm.

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Zhang *et al.*: Modes of vibration   2845

fold model was at about −4.5 mm and −3 mm in the superior-inferior direction for the aerodynamically driven mode and acoustically driven mode, respectively. For both modes of phonation, displacement amplitudes were greatest near the superior edge of the vocal fold. However, the vibration patterns were quite distinct. For the acoustically driven mode of phonation [Fig. 6(b)], a traveling wave was initiated about 10-mm below the superior edge, and traveled superiorly with time until it reached the superior edge of the vocal fold, where it exhibited its greatest amplitudes.

In contrast, for the aerodynamically driven mode of phonation [Fig. 6(a)], a forward or superiorly traveling wave and a backward or inferiorly traveling wave were initiated at about 5-mm below the superior surface. Although the backward-traveling wave had smaller amplitudes than the forward-traveling wave, it constituted a characteristic feature of the overall wave structure. The forward traveling wave proceeded until it reached the superior edge of the fold. The more complicated wave structure associated with the aerodynamic mode of phonation was further confirmed through means of a principal component analysis (PCA), as discussed below.

C. Empirical eigenfunctions

Empirical eigenfunctions (EEFs) were extracted from the three-dimensional displacement field of the entire medial surface of the vocal fold using the method of principal components analysis (PCA) (Berry *et al.*, 1994). Prior to analysis, the mean component of the displacement data was calculated and subtracted from the displacement data. Using the method of PCA, the three-dimensional displacement data over the medial surface $D(x,y,z,t)$ was decomposed into a linear combination of empirical eigenfunctions

$$D(x,y,z,t) = \sum_{k=1}^{N} \sigma_k \Phi_k(t) \Psi_k(x,y,z), \qquad (1)$$

where the normalized scalar function $\Phi_k(t)$ and normalized vector function $\Psi_k(x,y,z)$ were the kth temporal and spatial empirical eigenfunctions, σ_k was the modal scaling factor, and N was the total number of time samples. The temporal eigenfunctions may be also thought of as the temporal coefficients of the spatial eigenfunctions. The total energy associated with the vibration was given by

$$E_{tot} = \left\langle \int |D(x,y,z,t)|^2 dx dy dz \right\rangle_t = \sum_{k=1}^{N} \sigma_k^2, \qquad (2)$$

where the angle brackets denoted a time average. The relative weight of the contribution of each eigenfunction toward the total energy (or percentage variance) can be calculated as the ratio of corresponding σ_k^2 and the total energy E_{tot}.

Figure 7 shows the first two temporal eigenfunctions for both modes of phonation. The percentage variance (or relative weight) of the displacement data captured by the first four EEFs, and the vibration frequencies of the temporal eigenfunctions, are shown in Table I. The temporal eigenfunctions of the first four EEFs were harmonic functions with a dominant single-frequency component, with the first
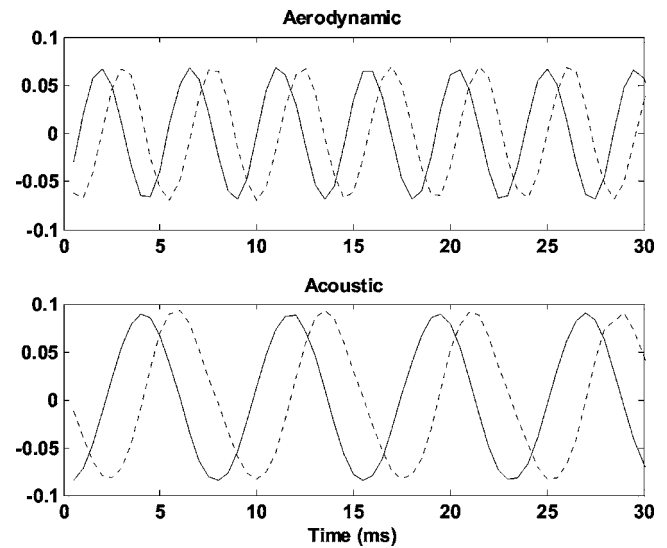
FIG. 7. The first two temporal empirical eigenfunctions for aerodynamically driven (top) and acoustically driven (bottom) modes of phonation. The tracheal tube length was 60 cm. —— : first temporal eigenfunction; – – – : second temporal eigenfunction.

two EEFs entrained to the fundamental frequency and the next two EEFs entrained to the second harmonic.

For both modes of phonation, Fig. 8 shows the normalized, medial-lateral component of the first two spatial eigenfunctions. From a coronal aspect, Fig. 9 illustrates medial-lateral and inferior-superior components of these same eigenfunctions, for one coronal slice of the medial surface, midway between anterior and posterior extremes. For the acoustically driven mode of phonation, considerable vertical motion was present (see Fig. 5), which was captured by the first eigenfunction (see Fig. 9). The medial-lateral component of the first eigenfunction was similar to an x-10 eigenmode (Titze and Strong, 1975; Berry *et al.*, 1994; Berry and Titze, 1996), in which the medial-lateral motion near the top of the glottal airway was in phase. This eigenfunction captured 85% of the variance of the medial surface displacements. Although the second eigenfunction was similar to an x-11 eigenmode and alternately shaped a convergent/divergent glottis (Titze and Strong, 1975; Berry *et al.*, 1994; Berry and Titze, 1996), this eigenfunction captured only 11.1% of the variance, and was approximately 7.6 times less energetic than the first eigenfunction. Other higher-order eigenfunctions captured 1.5% or less of the variance of the medial surface displacements. Therefore, in an acoustically driven mode of phonation, the vocal fold oscillations were captured primarily by one eigenfunction, with the medial-

TABLE I. Percentage variances (weights) and frequencies of the first four EEFs.

	Aerodynamic		Acoustic	
	Weight (%)	Frequency(Hz)	Weight (%)	Frequency(Hz)
EEF1	65.3	217	84.8	131
EEF2	31.3	217	11.1	131
EEF3	1.0	434	1.5	263
EEF4	0.5	434	1.1	263
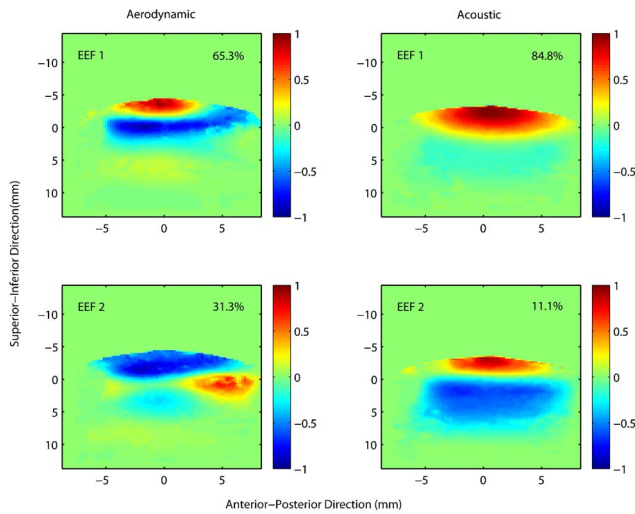
FIG. 8. (Color online) Normalized medial-lateral components of the first two spatial eigenfunctions for aerodynamically driven (left) and acoustically driven (right) modes of phonation. The tracheal tube length was 60 cm.

lateral component corresponding to an x-10 mode and the inferior-superior component corresponding to a z-10 mode.

In contrast, for the aerodynamically driven mode of phonation, the medial surface displacements were more complex, capturing two relatively strong eigenfunctions. The first eigenfunction (Figs. 8 and 9) was similar to an x-11 eigenmode (alternately shaping a convergent/divergent glottis) and captured 65% of the variance of the medial surface displacements. However, near the top of the glottal airway, the second eigenfunction was similar to an x-10 mode, and captured 31% of the variance of the medial surface displacements. Thus, for this aerodynamically driven mode of phonation, at least two eigenfunctions were required to reconstruct the primary dynamics of the vocal fold. Moreover, the two eigenfunctions entrained to the same frequency, which was independent of the subglottal acoustic resonance. This superposition of two eigenfunctions in the aerodynamically driven mode of phonation, with a phase difference determined by the temporal eigenfunctions, generated a traveling wave along the medial surface. This may explain the more complicated wave structure shown in Fig. 6(a), as compared to Fig. 6(b). Indeed, for the acoustically driven mode of phonation, only one eigenfunction was required to capture the essential dynamics of the vocal fold, resulting in a less-complicated wave structure, and a necessary dependence on and coupling with a nearby acoustic resonance.

IV. DISCUSSION

In this study, an attempt was made to distinguish aerodynamically and acoustically driven modes of vibration in a physical model based on the medial surface dynamics of the vocal fold, and the relative independence of the source and the acoustic resonator. The physical model was a single-layered, isotropic model of vocal fold vibration (Thomson *et al.*, 2005). Initially, only acoustically driven modes of phonation were observed in the physical model. For acoustically driven modes of phonation: (1) a single, dominant, x-10 like eigenmode captured the essential dynamics, explaining 85%
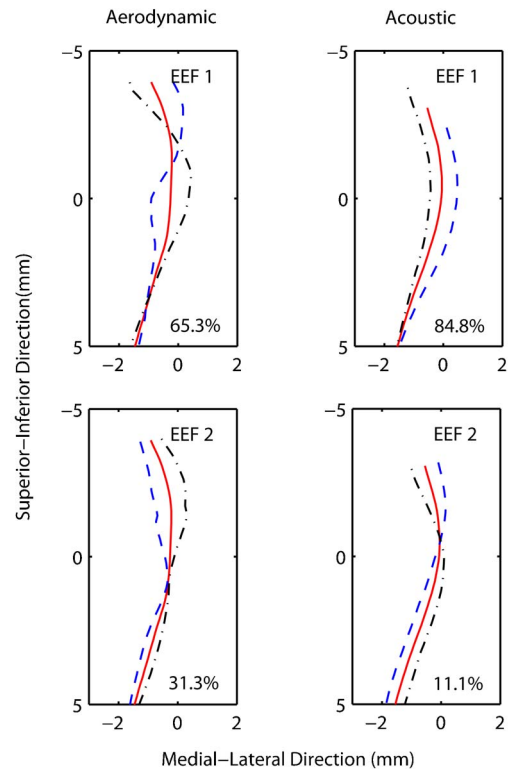
FIG. 9. (Color online) Medial-lateral and superior-inferior components of the first two spatial eigenfunctions for one coronal slice of the medial surface of the vocal fold, midway between anterior and posterior extremes, for both aerodynamically driven (left) and acoustically driven (right) modes of phonation. Maximum (dash-dotted) and minimum (dashed) projections of the eigenfunctions are superimposed on the mean surface projection (solid). The tracheal tube length was 60 cm.

of the variance of the medial surface displacements; (2) the eigenfunction always coupled with one of the acoustic resonances of the subglottal tract; and (3) phonation occurred only for relatively long tracheal lengths of 30 cm or more. In contrast, aerodynamically driven modes of phonation were observed only when a physical constraint restricted the vertical vibrations of the folds. For aerodynamically driven modes of phonation: (1) at least two eigenmodes were required to capture the essential dynamics, including an x-11-like eigenmode which explained 65% of the variance of the medial surface displacements, and an x-10-like eigenmode which explained 31% of the variance; (2) the two eigenmodes entrained to the same frequency and were successfully decoupled from subglottal acoustic resonances; and (3) self-oscillation was achieved for typical, human, tracheal lengths of approximately 17 cm.

The idea that human phonation, particularly the singing voice, may be influenced by sub- or supraglottal acoustic resonances is not new to speech science. Indeed, a variety of investigators have proposed an intimate connection between subglottal acoustic resonances and vocal registers (Nadoleczny-Millioud and Zimmerman, 1938; Vennard, 1967; van den Berg, 1968; Large, 1972; Austin and Titze, 1997). Similarly, Mergell and Herzel (1997) showed an example of high-frequency female singing, in which one vocal fold entrained with the first formant frequency, after the two folds had desynchronized due to left-right asymmetries.

Titze (1988) stated, "…the vocal instrument operates with varying degrees of source-resonator coupling. When coupling is weak…the larynx is able to control frequency, intensity, and the glottal source spectrum rather independently. Decoupling preserves constancy of phonation and seems to be desirable." The results of the present investigation demonstrate both weak and strong source-resonator coupling. The acoustically driven mode of phonation corresponds to strong source-resonator coupling, and the aerodynamically driven mode of phonation corresponds to weak source-resonator coupling. Although the physical model (Thomson *et al.*, 2005) had a propensity to vibrate in an acoustically driven mode of phonation, we were able to induce an aerodynamically driven mode of phonation by applying an appropriate vertical restraint. In the normal human condition, epilaryngeal manipulation through means of the ventricular folds may be able to provide a natural vertical restraint on the true vocal folds. Simultaneous contraction and stiffening of the thyroarytenoid muscle may also cause a natural vertical restraint.

In retrospect, after performing the experiment, it was realized the Young's modulus of 11 kPa utilized for the vocal fold model was relatively stiff, especially for the vocal fold cover, which may have discouraged the propagation of the mucosal wave. By lowering the value of the Young's modulus, post-experiment studies revealed a whole continuum of source-resonator coupling possibilities, with the examples in this investigation representing the extremes of both strong source-resonator coupling (e.g., acoustically driven modes of phonation) and weak source-resonator coupling (e.g., aerodynamically driven modes of phonation). Further details regarding the continuum of source-resonator coupling possibilities will be the subject of future work.

While the effectiveness of the vertical constraint in facilitating aerodynamically driven modes of phonation is not fully understood, it has potential applications in the area of phonosurgery which might merit further investigation. For example, despite the existence of a relatively stiff mucosal cover in the current study (both body and cover had a Young's modulus of approximately 11 kPa), a vertical constraint facilitated the emergence of an x-11 eigenmode, and the propagation of a mucosal wave. Whether induced by aging, scarring, or some other means, a stiffened mucosal cover is clinically known to be a major deterrent to mucosal wave propagation. Thus, it may be that the introduction of a vertical constraint through phonosurgical means could have potential benefit for patients with a stiffened vocal fold cover. Similarly, the introduction of a vertical constraint may also facilitate the propagation of a mucosal wave in patients with a relatively lax vocal fold body, such as might be caused by RLN (recurrent laryngeal nerve) paralysis.

Although no supraglottal vocal tracts were included in the current experiments, we would expect the supraglottal vocal tract to exhibit effects similar to that of the subglottal system, especially near phonation onset. Moreover, the inclusion of a supraglottal vocal tract would introduce yet another resonator, further complicating the interactions of the coupled system. This topic will be explored in future studies.

While medial surface dynamics were a useful tool to distinguish aerodynamically and acoustically driven modes of phonation in this study, such dynamics may not be a useful tool to distinguish these modes of phonation in other contexts, since medial surface dynamics are rarely reported. A few exceptions are Berry *et al.* (1994), Berry *et al.* (2001), and Doellinger *et al.*, (2005a, 2005b). In nearly all of these studies, the x-11 eigenmode was the strongest EEF, as might be expected in an aerodynamically driven mode of phonation. However, in Berry *et al.* (2001), the x-10 eigenmode was stronger, which was indicative of strong source-resonator coupling. Indeed, in a related conference proceeding, Berry *et al.* (1999) indicated that they did observe involuntary frequency jumps, which may be another indication of source-tract coupling.

Indeed, a variety of commonly observed vocal phenomena may help to distinguish the degree of source-resonator coupling when access to medial surface dynamics is not available. For example, continuous control of fundamental frequency as a function of vocal fold tension or subglottal pressure may be suggestive of aerodynamic phonation (e.g., independence of source and resonator). On the other hand, involuntary frequency jumps or "voice breaks" may be suggestive of source-resonator coupling. Because such jumps may also be caused by the intrinsic dynamics of the vocal folds (e.g., different eigenmodes within the folds may be excited), voice breaks do not guarantee the existence of source-tract coupling. However, when voice breaks occur, source-tract coupling is highly probable, and must not be ruled out.

V. CONCLUSION

Using a physical model of vocal fold vibration, aerodynamically and acoustically driven modes of vibration were distinguished based on the medial surface dynamics of the vocal fold, and the relative independence of source and acoustic resonator. The single-layered, cover-only, isotropic physical model was dynamically similar to a one-mass model, and its medial-lateral vibration could be described efficiently by an x-10 eigenmode. The physical model had a propensity to vibrate in an acoustically driven mode of phonation, i.e., vocal fold vibration did not occur unless it was coupled to an acoustic resonance of the vocal system. However, an aerodynamically driven mode of phonation could be facilitated by introducing a vertical constraint, and or by lowering the stiffness of the vocal fold. This mode of phonation captured two relatively strong eigenmodes (e.g., x-11 and x-10), exhibited a clear mucosal wave, and was vibrated independently of acoustic resonances. Because of the increased control over fundamental frequency and register, aerodynamically driven modes of phonation are generally deemed preferable to acoustically driven modes of phonation in both speech and singing.

ACKNOWLEDGMENTS

Adachi, S., and Sato, M. (**1996**). "Trumpet sound simulation using a two-dimensional lip vibration model," J. Acoust. Soc. Am. **99**(2), 1200–1209.

Austin, S. F., and Titze, I. R. (**1997**). "The effect of subglottal resonance upon vocal fold vibration," J. Voice **11**, 391–402.

Berry, D. A. (**2001**). "Mechanisms of modal and nonmodal phonation," J. Phonetics **29**, 431–450.

Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604.

Berry, D. A., Montequin, D. W., and Tayama, N. (**2001**). "High-speed digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547.

Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345.

Berry, D. A., Titze, I. R., and Herzel, H. (**1999**). "Empirical eigenfunctions obtained from highspeed imaging of the vocal folds," J. Acoust. Soc. Am. **105**(2), 1304.

Broad, D. (**1979**). "The new theories of vocal fold vibration," in *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass (Academic, New York).

Chan, R. W., Titze, I. R., and Titze, M. R. (**1997**). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727.

Copley, D. C., and Strong, W. J. (**1996**). "A stroboscopic study of lip vibrations in a trombone," J. Acoust. Soc. Am. **99**(2), 1219–1226.

Doellinger, M., Berry, D. A., and Berke, G. S. (**2005a**). "Medial surface dynamics of an *in vivo* canine vocal fold during phonation," J. Acoust. Soc. Am. **117**, 3174–3183.

Doellinger, M., Tayama, N., and Berry, D. A. (**2005b**). "Empirical eigenfunctions and medial surface dynamics of a human vocal fold," Methods Inf. Med. **44**, 384–391.

Flanagan, J. L. (**1958**). "Some properties of the glottal sound source," J. Speech Hear. Res. **1**, 99–116.

Flanagan, J. L., and Landgraf, L. (**1968**). "Self-oscillating source for vocal-tract synthesizers," IEEE Trans. Audio Electroacoust. **AU-16**, 57–64.

Fletcher, N. H. (**1993**). "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180.

Hartley, R., and Zisserman, A. (**2000**). *Multiple View Geometry in Computer Vision* (Cambridge University Press, UK).

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.

Ishizaka, K., and Matsudaira, M. (**1972**). "Fluid mechanical considerations of Vocal Cord Vibration," Monogra. 8, Speech Commun. Res. Lab., Santa Barbara, CA.

Ishizaka, K., Matsudaira, M., and Kaneko, T. (**1976**). "Input acoustic-impedance measurement of the subglottal system," J. Acoust. Soc. Am. **60**, 190–197.

Jiang, J. J., and Titze, I. R. (**1993**). "A methodological study of hemilaryngeal phonation," Laryngoscope **103**, 872–882.

Large, J. (**1972**). "Towards an integrated physiologic-acoustic theory of vocal registers," NATS Bull. **29**, 18–25.

Mergell, P., and Herzel, H. (**1997**). "Modelling biphonation - The role of the vocal tract," Speech Commun. **22**, 141–154.

Nadoleczny-Millioud, M., and Zimmerman, R. (**1938**). "Categories et registres de la voix," Names Registers Voice **23**, 21–31.

Stevens, K. N. (**1977**). "Physics of laryngeal behavior and larynx modes," Phonetica **34**, 264–279.

Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.

Titze, I. R., and Strong, W. J. (**1975**). "Normal modes in vocal fold tissues," J. Acoust. Soc. Am. **57**, 736–744.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**(4), 1536–1552.

Titze, I. R., Schmidt, S. S., and Titze, M. R. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084.

van den Berg, J. W. (**1968**). "Register problems," Ann. N.Y. Acad. Sci. **155**, 129–135.

Vennard, W. (**1967**). *Singing … the Mechanisms and the Technique*, revised ed. 2 (Carl Fisher, Inc., New York).

Zhang, Z., Neubauer, J., and Berry, D. A. (**2006**). "The influence of subglottal acoustics on laboratory models of phonation," J. Acoust. Soc. Am. **120** (3), 1558–1569.

Perception of synthetic vowel exemplars of 4 year old children and estimation of their corresponding vocal tract shapes

Richard S. McGowan

CReSS LLC and Haskins Laboratories, 1 Seaborn Place, Lexington, Massachusetts 02420

Formant scalings for vowel exemplars of American 4 year olds who were imitating adult production were used along with published data of American adult male vowel production to synthesize /ɑ, æ, u, i/. Other vowel exemplars were also synthesized. Adult listeners were asked to categorize these synthetic vowels in a forced choice task. With some exceptions, the formant frequencies preferred for the vowels /ɑ, æ, u, i/ were close to the published data. In order to gain insight on children's articulation during imitation of vowels /ɑ, æ, u, i/, a five-tube model was used in an algorithm to infer vocal tract shape from the first three formant frequencies of the adult productions, the formant frequencies derived for 4 year olds by scaling, and formant frequencies for 4 year olds derived based on the listening experiments. It was found that the rear tube length for the children, in proportionate terms, was nearly always greater than that of the adult. The rear tube length was proportionately twice as long in children compared to adults for the vowel /u/. Tongue root flexibility and the oblique angle between the pharynx and mouth may be more important than pharynx length in determining formant scalings for 4 year old children. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2345833]

I. INTRODUCTION

Children generally learn to speak the language that is used in their surroundings. Within the first 2 years, children will have acquired a basic vocabulary that is understandable to many adults in their families. However, the acoustic output of young children's speech cannot completely resemble adults' speech simply because of anatomical differences between adults and young children. One of the most obvious of these differences between children and adults is in vocal tract length, but there are also differences in the relative lengths of the pharynx and mouth (Goldstein, 1980; Kent and Vorperian, 1995; Fitch and Giedd, 1999). Despite anatomical differences, children do learn to speak, and, in particular, they produce vowels that are recognized by adults.

Acoustic studies of vowel production by American men, women, and/or children, such as those of Peterson and Barney (1952), Hillenbrand, Getty, Clark, and Wheeler (1995), and Lee, Potamianos, and Narayanan (1999) have documented the acoustics of vowel production differences between adults and children. A quantitative look at such data indicates that there is more than simple length scaling that must be invoked to account for formant frequency differences between, say, men and children (e.g., Fant, 1975). Some of the differences might be attributed to differences in the pharynx-to-mouth length differences. However, the effect of anatomical differences on articulation is still unclear.

One way to understand the relation between adults' production and children's production of particular utterances is to examine the children's production when they are asked to imitate the speech of adults. Imitation is understood here in the narrow sense of the situation when children are asked to imitate adult phonetic segments in experimental context. Here, we will restrict ourselves to vowel production and ask

the following questions. How do children imitate adult vowels when their anatomies differ? Are there any constraints on children's production of vowels beyond the length differences in fixed vocal tract structures? The answers to these questions will help in understanding how children develop as speakers.

One source of data on children's vowel production when they were asked to imitate adult vowels comes from Kent and Forner (1979). In their experiment, Kent and Forner asked three 4 year old boys and six 4 year old girls to imitate vowels that had been synthesized on a cascade formant (Klatt) synthesizer. Five of these synthetic vowels were modeled on the mean adult male data of Peterson and Barney (1952) for the vowels /i, u, ɑ, æ, ɚ/, and five other vowels were used to fill the vowel space, but they were without specific phonemic identity. These ten stimuli were presented five times each for a total of 50 imitations. Children's formant frequencies from the imitation task were made from spectrograms in a procedure described in their paper (Kent and Forner, 1979). From these measures, an average scale factor (=ratio of child formant frequency to adult formant frequency) for each formant frequency and each vowel can be computed across all ten 4 year olds (see Fig. 10 of Kent and Forner, 1979 for a closely related quantity). Based on these formant scaling factors for 4 year old children and the formant frequencies for adult males provided by Olive, Greenwood, and Coleman (1993), a set of formant frequencies for vowels /ɑ, æ, u, i/ as might be spoken by a 4 year old imitating an adult can be found. It should be noted that the scaling derived from Kent and Forner and the formant frequencies provided by Olive *et al.* (1993) could be affected differently by individual and regional dialect differences.

Listening experiments with adult subjects were used to verify that the formant values for the four year-old children's

vowels derived using the Kent and Forner (1979) and Olive *et al.* (1993) parameters were, indeed, reasonable. Thus, using these data, vowels /ɑ, æ, u, i/ were synthesized, as well as vowels whose formant frequencies were altered from these four vowels. This also served to verify the scaling results of Kent and Forner (1979) when they are applied to a set of published adult formant frequency values of Olive *et al.* (1993).

In order to begin to answer the questions about how children imitate it is necessary to know something about the articulation employed by children to imitate the vowels of adult speech. It is difficult to obtain articulatory measures during speech production in young children. Therefore, our strategy was to infer vocal tract shape from the speech acoustic data of children's vowel production described above. The vocal tract shapes of both children and adult males in the production of vowels /ɑ, æ, u, i/ were inferred using an analysis-by-synthesis procedure that maps formant frequencies to an acoustic tube model of the vocal tract. While vocal tract shapes are not synonymous with articulatory positions, vocal tract shape can be informative as to the articulatory strategy used by children during speech imitation.

Our previous work in inferring vocal tract shape and dimensions from acoustics has been to use analysis-by-synthesis, in which the synthesis is done using an articulatory synthesizer with a model vocal tract. In this previous work, task-dynamic parameters of the model vocal tract were adjusted in a stochastic optimization algorithm, a genetic algorithm, so that the resulting formant frequency trajectories produced by an articulatory synthesizer matched the formant frequencies of the corresponding data vowel (e.g., McGowan, 1994). The closeness of the match for each utterance was judged by the size of sum of the square differences between the model and the data in the first three formant frequencies. The work here will use static formant frequency values to infer static vocal tract tube shapes. There can be problems of ambiguity (i.e., multiple optimal solutions) in such a procedure, so the analysis-by-synthesis procedure was run several times, and, further, the number of allowed vocal tract tube sections was limited. By running a stochastic optimization procedure several times, it can be expected that a variety of optimal solutions will be found. Also, by attending to grosser features of the vocal tract corresponding to the information that the first three formant frequencies provide, it can be assured that the inferences made about the vocal tract shapes are valid.

II. LISTENING EXPERIMENT: METHOD

Vowels /ɑ, æ, u, i/ were synthesized as monophthongal vowels with a cascade (Klatt) formant synthesizer using the formant frequencies shown in Table I. The first three formant frequencies (F1, F2, F3) were derived from the scaling factors for 4 year olds provided by Kent and Forner (1979, Fig. 10) and the formant frequencies for an adult male provided by Olive *et al.* (1993, p. 104). The formant frequencies were measured from sentences containing the words "bottle," "bat," "boot," and "beet" spoken by an adult male from Pittsburgh, Pennsylvania (Olive *et al.*, 1993, p. 8). The formant

TABLE I. The first three formant frequencies, without alteration, used to synthesize KFO condition vowels of a 4 year old. The scaling factors from Kent and Forner (1979) and adult formant frequencies from Olive *et al.* (1993) are in brackets.

	Child F1 (scale, Adult F1)	Child F2 (scale, Adult F2)	Child F3 (scale, Adult F3)
ɑ	1125 Hz (1.5, 750 Hz)	1650 Hz (1.5, 1100 Hz)	4030 Hz (1.55, 2600 Hz)
æ	980 Hz (1.4, 700 Hz)	2475 Hz (1.5, 1650 Hz)	4250 Hz (1.7, 2500 Hz)
u	429 Hz (1.43, 300 Hz)	1080 Hz (1.27, 850 Hz)	4375 Hz (1.75, 2240 Hz)
i	448 Hz (1.6, 280 Hz)	3488 Hz (1.55, 2250 Hz)	4125 Hz (1.5, 2750 Hz)

frequencies from Olive *et al.* (1993) are all within 10% of the mean values for adult males given in Peterson and Barney (1952), which were the values used by Kent and Forner (1979) in their experiments. Further, with the exception of the three formants for /u/ and F2 for /ɑ/, the formant frequencies are within one standard deviation of the mean values given by Lee *et al.* (1999). In the cases of these four formant frequency values, the values of Olive *et al.* (1993) are within 20 Hz of the values given by Peterson and Barney (1952). Because the speech data for the Lee *et al.* (1999) were collected in St. Louis, Missouri and the data of Peterson and Barney (1952) were collected in New Jersey over 40 years previously, these differences can probably be attributed to dialect differences amongst the various studies. With the values of F1, F2, and F3 for the synthetic children's vowels derived from Kent and Forner and from Olive *et al.*, the values F4 and F5 were scaled from 3500 Hz and 4500 Hz using the scaling factor for F3. This was done because of a lack of data for F4 and F5. These synthetic base vowels, /ɑ, æ, u, i/, are called *KFO condition* vowels in this paper. These vowels were intended to be synthetic approximations of vowels produced by 4 year olds imitating adults.

Other vowels were synthesized with the first three formant frequencies of the base vowels altered by −20%, 0%, or 20%, and the fourth and fifth formant frequencies were changed the same amount as the third formant frequency from the values for the base vowels. With all possible combinations of these perturbations there were twenty-seven $=3^3$ different vowel exemplars for each base vowel, including the base vowel itself. All vowels were sampled at a 20 kHz sampling rate. The vowels were 400 ms long with the fundamental frequency gradually decreasing from 240 Hz to 200 Hz throughout that duration. This was on the low side for 4 year old fundamental frequencies [e.g., Eguchi and Hirsh (1969) show a mean fundamental frequency of 286 Hz], but they were within reason and engendered a less ambiguous vowel quality than would have higher fundamental frequencies. Also, after an initial increase in amplitude for the first 20 ms, the voice amplitude decreased while parameters making the voice sound more breathy, such as random noise amplitude, increased with time into the vowel. These parameter settings made the voice quality child-like, as judged by the author.

TABLE II. The mean of the formant frequencies in each vowel category receiving a score of 50%, or greater, of the highest score. The scaling factor from adult male formants and percent difference from the children's KFO condition formant frequencies are shown in Table I in brackets.

	Child F1 (scaling, % diff.)	Child F2 (scaling, % diff.)	Child F3 (scaling, % diff.)
ɑ	1210 Hz (1.61, +7.5%)	1672 Hz (1.52, +1%)	4128 Hz (1.59, +2%)
æ	1087 (1.55, +11%)	2700 (1.64, +9%)	4173 (1.67, −2%)
u	393 (1.31, −8%)	1080 (1.27, 0%)	4421 (1.97, +1%)
i	386 (1.38, −14%)	3541 (1.57, +2%)	4188 (1.52, +2%)

Fifteen adult students who were recruited by advertisement from the University of Connecticut student body participated in a listening experiment with these synthetic vowels. Each subject was seated at a computer terminal and listened to the stimuli over a set of speakers controlled by the computer. The 108 synthetic vowels were presented in randomized lists, in which each token appeared three times. As soon as a stimulus was played, a list of five hVd words (except for one hVti word) appeared on the screen (see the Appendix). The subject was asked to choose the word with the vowel that most closely matched the vowel presented auditorily. Once the choice was made, the subject was asked for the goodness of the match on a scale from one to five. After the subject responded to this choice, he was given the choice to repeat the trial, continue on to the next stimulus, or to exit the experiment. In this way the subject was able to be as certain as possible about his judgments and less likely to attempt to compensate for a perceived poor judgment in listening to a future stimulus.

III. LISTENING EXPERIMENT: RESULTS

For each synthetic vowel, the goodness scores for each category were totaled. For instance, a certain synthetic vowel would sometimes be heard as /i/ and sometimes as /ɪ/, and the goodness scores for this vowel were totaled separately for each of these phonemic categories. Table II shows the average formant frequencies of the synthetic vowels that received scores of 50% (an arbitrary level), or greater, of the maximum goodness score for the phonemic categories of each of the four base vowels. The percents in brackets are the amount these formant frequencies differ from those in Table I. These average formant frequency values specify vowels that are called *listening condition* vowels in this paper.

With some exceptions, listeners favored the formant frequency values for the 4 year olds that were taken to be representative of children's productions according to the Kent and Forner (1979) scalings and the Olive *et al.* (1993) formant frequencies. The percents in brackets indicate that listeners favored more extreme F1 values, or extreme tongue height positions, than those given in the KFO condition. Further, the listeners favored a higher F2, or more fronted tongue, for /æ/.

TABLE III. The number of times a particular vowel, as determined by the KFO condition, is classified into one of five vowel categories with a goodness score of greater than, or equal, to 3.

KFO condition vowel	scores	Listener Choices				
		ɑ	ɔ	ʌ	æ	oʊ
ɑ	**5, 4**	9	7	1	0	0
ɑ	**3**	9	6	2	3	2
		æ	ɛ	ɑ	eɪ	ʌ
æ	**5, 4**	11	4	0	0	0
æ	**3**	0	8	1	0	0
		u	ʊ	oʊ	ʌ	ɑ
u	**5, 4**	10	7	1	0	0
u	**3**	7	2	2	2	0
		i	ɪ	eɪ	ɛ	æ
i	**5, 4**	20	1	2	0	0
i	**3**	11	0	1	1	0

The KFO condition vowels, whose formants are shown in Table I, were examined to find which categories listeners favored. The instances when these vowels received a goodness score of four or five for a particular category were counted, and the instances when these vowels received a goodness score of three were counted separately. Dividing the goodness scores four and five from goodness score three indicates how many times a token was rated as very good or excellent as an exemplar from when it was merely rated as a moderately good exemplar. The total number of times that a KFO condition vowel was rated is 45 (=15 subjects times 3 repetitions). Table III shows these counts for each KFO condition vowel.

Table III accounts for 39 tokens of /ɑ/, 24 tokens of /æ/, 31 tokens of /u/, and 36 tokens /i/, out of 45 tokens each. As indicated by Table III, the four vowels in the KFO condition were acceptable tokens of their target phonemic category in the listening condition. The low vowels /ɑ/ and /æ/ in the KFO condition were often perceived as good examples of slightly higher, neighboring phonemes, that is as /ɔ/ and /ɛ/, respectively. This is consistent with Table II, which indicates that listeners preferred higher first formant frequencies (i.e., a lower vowel) than provided by these vowels in the KFO condition. Also, the KFO condition /u/ was often heard as a good example of /ʊ/, a neighboring lower vowel. This, again, is consistent with Table II: listeners preferred higher vowels (i.e., lower first formant frequencies) for /u/ than provided in the KFO condition. While listeners preferred higher vowels than provided by the KFO condition /i/, they still counted the KFO condition /i/ as an unambiguously good example of that vowel.

IV. VOCAL TRACT DIMENSIONS: METHOD

Children's area functions for the vowels /ɑ, æ, u, i/ were inferred from formant frequency data from two different sources: the KFO condition for 4 year old children (Table I)

and preferences given by adults in the listening experiment (Table II). A five-tube vocal tract model was used, which is justified by the fact that the four adult male vowels could be simulated using as few as four tube sections (Stevens, 1998), thus leaving one extra tube in case the children's vocal tract shapes proved anomalous. Adult male area functions were also inferred for these vowels.

A genetic algorithm was used for each vowel to infer the lengths and areas of the five tubes that could produce the formant frequencies given in Tables I and II. A simple genetic algorithm was used as in previous work in recovering articulatory movement from formant frequencies (e.g., McGowan, 1994), but only static area functions were inferred in the current work. A genetic algorithm is a stochastic optimization technique, in which each variable parameter is coded as a binary number, which is called a gene (Goldberg, 1989). All the genes, each coding a particular parameter, are concatenated to form a chromosome. In this study a gene represented either a tube length or a tube area, for a total of ten genes in a chromosome. For the present study, the fitness of a chromosome was based on the first three formant frequencies produced by the five-tube model specified by the genes of that chromosome. Those three formant frequencies were compared with formant frequency data from one of the conditions in Table I and II and a distance between the model formants and data formants was computed. This distance was related to fitness, so that the smaller the distance, the greater the fitness. The specific form of the relation between distance and fitness is given below.

A simple genetic algorithm starts with a random group, or population, of chromosomes, each coding a different set of parameters. The population is then allowed to progress through generations. For each generation, pairs of chromosomes are selected for possible mating. In the algorithm used in this study, the number of pairs equaled one-half the size of the population. The chromosomes for possible mating are selected according to their fitness, so that the probability of a chromosome being selected in this work was proportionate to its fitness. Each pair of chromosomes selected has a probability of mating. If they do not mate, they are put back into the population unchanged, except for a small probability of mutation, where bits of the chromosome are changed in a process of mutation. Otherwise each of the chromosomes is cut at a randomly selected bit in each chromosome, and the divided strings of bits are swapped to produce new offspring chromosomes. These chromosomes are then subjected to the same small probability of mutation as the unmated pairs of chromosomes. The fitness of each new chromosome is evaluated, and a new generation is thus created. In the algorithm used here, the fittest individual of a generation was always preserved to enter the next generation. The process is stopped after a fixed number of generations. This process is supposed to be analogous to natural selection, and it can be shown to tend to produce chromosomes that are more fit as the number of generations increase (Goldberg, 1989).

In the current study, populations of 240 chromosomes were coded with five bits per tube area with minimum areas of 0.05 cm² and maximum areas of 10 cm², and with five bits per tube length with minimum lengths of 0.5 cm and

maximum lengths of 8.0 cm. These populations were run through 120 generations with fitness proportionate selection. There was a 0.6 chance for two selected chromosomes to mate, where mating entailed cutting a chromosome at a random bit and swapping the cut substrings to produce children chromosomes. There was a 0.005 chance for any bit to mutate from a one to a zero or vice versa. To compute the fitness function, the square-root of the sum over the first three formant frequencies of the square fraction of the differences between the data formant frequency and the inferred formant frequency was computed. The fitness was the exponential of the inverse of this quantity. The genetic algorithms were run five times for each of the sets of three formant frequencies for the four vowels from both Tables I and II for the children, and from Table I for adult males.

V. VOCAL TRACT DIMENSIONS: RESULTS

The vocal tract dimensions that best fit the adult male acoustic data are shown in Fig. 1. For the adults there were often multiple recovered area functions. Also, while the cross-sectional area near the glottis varies from less than 0.5 cm² for /ɑ/ [Fig. 1(a)] to over 3.0 cm² for /i/ [Fig. 1(d)], these areas only represent an average cross-sectional area in the pharyngeal region. For the adult /i/ [Fig. 1(d)], there is some ambiguity in the area of the tube in front of the constriction, because there are compensatory area changes in the rear cavity. For the adult /ɑ/ [Fig. 1(a)] there is a compensatory relation between the length of the rear tube and the total vocal tract length. The most striking ambiguity is for the vowel /æ/ [Fig. 1(b)], where, in the first case, there is a narrow rear tube for about one-third of the total length of the vocal tract with a wide front tube, and in the other case there is a more gradual increase in area from the rear to the front of the vocal tract. Figure 2 shows the spectra for the two adult /æ/s. Because the second formant has most of its energy associated with the longer front tube in the first case, it has greater amplitude than the third formant, which is associated with the rear tube [Fig. 2(a)]. In the second case the third formant has greater amplitude than the second formant when there is a more gradual increase in area [Fig. 2(b)]. Thus more of the second formant energy is associated with the rear and more of the third formant energy is associated with the front of the vocal tract compared to the first case. Accounting for relative formant amplitudes would disambiguate these situations.

Comparisons of the tube shapes from the adults' production, the children's KFO condition, and the adult listening preferences of children's synthetic vowels show the same general vocal tract shapes (Figs. 1, 3, and 4). Both /æ/ and /ɑ/ have rear constriction tubes, with the one for /ɑ/ more constricted than the one for /æ/ [Figs. 1(a), 1(b), 3(a), 3(b), 4(a), and 4(b)]. In both the KFO and listening conditions, the children's /æ/ had a less, constricted, but longer rear tube than the adult case with a rear tube length about one-third the vocal tract length. The children's /æ/ is more closely related to the second adult case with a more gradually changing area function. However, there were configurations with more constricted, shorter rear tubes for the children's KFO and listen-
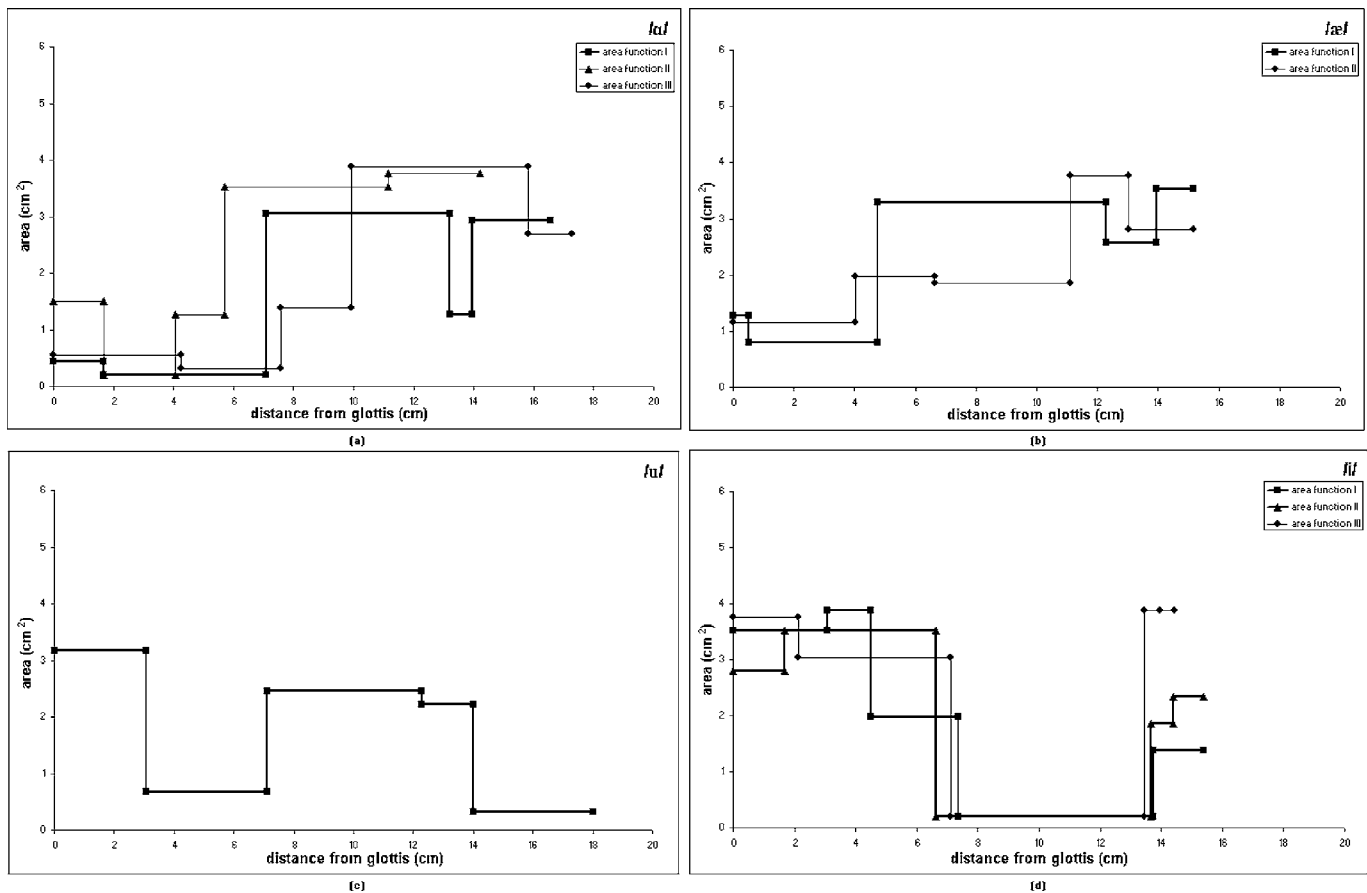
FIG. 1. Tube dimensions recovered from adult formant frequency data (a) /ɑ/, (b) /æ/, (c) /u/, and (d) /i/. Different recovered vocal tracts for a single vowel are possible and are denoted by different symbols and line thickness.

ing conditions that did not attain the fitness of the configurations shown in Figs. 3(b) and 4(b), but were nearby in terms of formant frequency fit. Thus there appears to be a similar ambiguity to the adult males in the vowel /æ/ for children. For /u/ there are two cavities separated by a constriction, and there is a constriction at the lips [Figs. 1(c), 3(c), and 4(c)]. For /i/ there is a back cavity with a front constriction [Figs. 1(d), 3(d), and 4(d)].

Beyond the general shapes of the vocal tract tubes there were differences between the children's and adult's vocal tract tubes. A difference that appeared fairly consistently across the four vowels involves the length dimension of the rear tube, whether or not that tube was a constriction or a cavity[1]. Comparisons of these lengths and their ratios with the total recovered vocal tract lengths are shown in Table IV. (Recall that there were multiple tube configurations for the adult data, as shown in Fig. 1.)

All of the rear tube length ratios in the case of the children's vowels in the listening condition were at least as large as the corresponding adult rear tube length ratios. The same is true for the children's vowels, except for /i/, in the KFO condition. Even for /i/ in the KFO condition, the adults' ratios do not exceed the children's ratios by more than 0.06. In the case of /u/, the ratio is on the order twice as large in children compared to the adults. /u/ is the vowel that has the greatest range of the three formant scalings: 1.27–1.75 for the KFO condition (Table I) and 1.27–1.97 for the listening

condition (Table II). For the KFO condition of /u/, the scaling difference was 1.43 for F1 and 1.27 for F2, which is the greatest difference among the vowels for these two formants. Thus, it is not surprising to find substantial differences in the ratios of tube lengths between adults and children. The recovery algorithm indicates that on a proportionate length basis, the children's /u/s are more "fronted" than the adult /u/. Also, this is the vowel of the four with the shortest rear tube length for both children and adults.

It is interesting to compare the rear tube lengths to the location of the velum. In one method of computation using MRI images, Fitch and Giedd (1999) took the distance from the glottis to the free edge of the velum (uvula) as the length of each of their subject's pharynx. The pharyngeal segment, along with four other line segments spanning the rest of the upper vocal tract, was also used to compute the total length of each vocal tract in their study. From their data, the ratio of a 4 year old's pharynx length to total length is between 0.25 and 0.29. For an adult male, this ratio is 0.37. In reference to Table IV, this means that all the children's rear tubes include the uvula, although for /u/ the uvula is very close to the front of the rear tube. For the adult male however, it is possible, according to these analyses, that the rear tube does not include the uvula for /ɑ/ and /æ/, and is wholly contained in the pharynx below the uvula. Further, the rear tube for the adult male /u/ is wholly contained in the pharynx below the uvula.
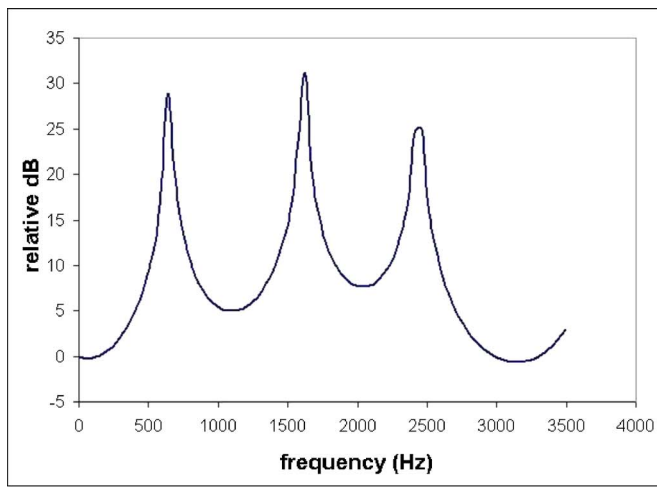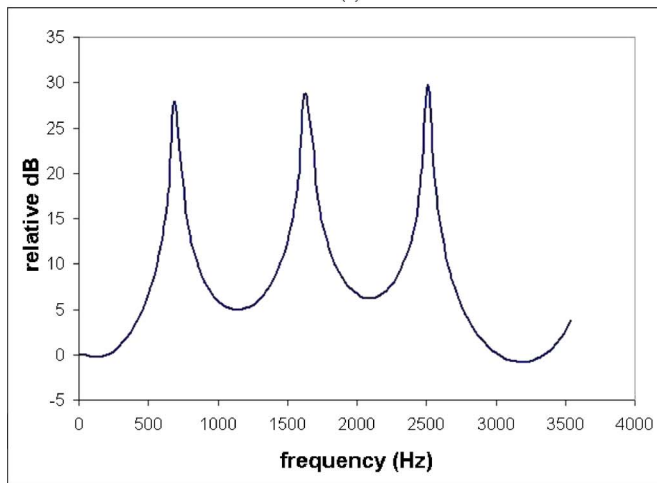
FIG. 2. Transfer functions for two tube configurations for adult /ae/ (a) short rear tube, and (b) long rear tube.

Because the rear tube in /u/ is a part of the volume element in one of the Helmholtz resonators in a double Helmholtz resonator system (Fant, 1960), it is possible to explore the acoustic ramifications of this geometry. F1 and F2 in /u/ are the resonances of this double Helmholtz resonance system, with the rear tube and the constriction closest to the rear constituting one resonator, the back resonator, and the mouth cavity and lip constriction forming the other Helmholtz resonator, the front resonator. The resonant frequencies of these resonators, if they were uncoupled are denoted $F1_0$ and $F2_0$ here, with the $F1_0$ the lower frequency. When these Helmholtz resonances are coupled, as they are in the vocal tract, the resonance frequencies shift from their uncoupled values to F1 and F2, with $F1 < F1_0$ and $F2 > F2_0$. Estimates of the resonant frequencies of the uncoupled back and front resonators are shown in Table V.

It can be seen that $F1_0$ and $F2_0$ are closer together for both of the children's conditions than for the adult's, both in absolute and relative terms. This corresponds to the fact that in both the production and listening condition for children, the scaling factor for F1 is greater than that for F2, which means that F1 is relatively closer to F2 for both children's conditions than for the adult male. (The proximity of F1 to

F2 is not completely accounted for by the proximity of $F1_0$ and $F2_0$, but also depends on the degree of coupling between the resonators.) Further, while the lower frequency $F1_0$ is associated with the front resonator for the adult and the listening condition for the children, it is associated with the back resonator in the KFO condition for the children.

Based on these results, 4 year old children are able to form vocal tract shapes that are appropriate for the vowels /ɑ, æ, u, i/. However, not all of these vowels are related to the adult versions of these vowels by a simple formant scaling factor. This is most apparent with /u/, and to some extent /æ/, in the data from Kent and Forner (1979). This acoustic pattern appears to be the result of a proportionately longer rear cavity for /u/, and, possibly, a longer rear tube for /æ/ in the child's vocal tract, compared to adults.

VI. SUMMARY AND DISCUSSION

The adult formant frequencies for the vowels /ɑ, æ, u, i/ as given by Olive *et al.* (1993), were scaled to the formant frequencies of 4 year old children imitating adults according to the values given by Kent and Forner (1979). These KFO condition vowels, along with vowels with formant frequencies altered from these values, were synthesized on a cascade formant synthesizer employing a child-like voice quality. Adults were asked to categorize and then to assign goodness scores to these vowels in a forced choice listening task. The average of each formant frequency for each vowel that received at least 50% of the score of the vowel with the largest total score in each phonemic category was computed. The results for the four vowels /ɑ, æ, u, i/ indicated that the formant frequencies of vowel tokens preferred by adult listeners differ little from those computed from adult data and scale factors. The largest differences between the formants in the listening condition and the KFO condition formants were more extreme F1s for all four vowels and the higher F2 for /æ/ in the listening condition. However, the KFO condition vowels, /ɑ, æ, u, i/, were usually perceived as good exemplars of their vowel categories in the listening condition.

Five-tube models were used in an analysis-by-synthesis procedure to infer the lengths and cross-sectional areas of the tubes from the first three formant frequencies. There was some minor ambiguity in the optimum five tubes for the adult male /ɑ/ and /i/. However, there was a more striking two-way ambiguity in the adult male /æ/, which could be differentiated if formant amplitudes had been accounted for along with formant frequencies. While adult males may produce /æ/ in a configuration with a relatively short and constricted rear tube or with an alternative tube with a gradually changing area, the inferred shapes for children show that they may prefer the latter configuration. Informal experimentation shows ambiguity for the mid, front vowel /ɛ/ as well. These observations on low-to-mid front vowels will be pursued in subsequent work, noting here that they are vowels with difficult-to-define constriction locations (e.g., Ladefoged and Maddieson, 1996; p. 284).

In both the listening and KFO conditions, children appear to produce back vowels /ɑ/ and /u/ with a rear tube at least as long, proportionately, as adults. (The rear tube is a
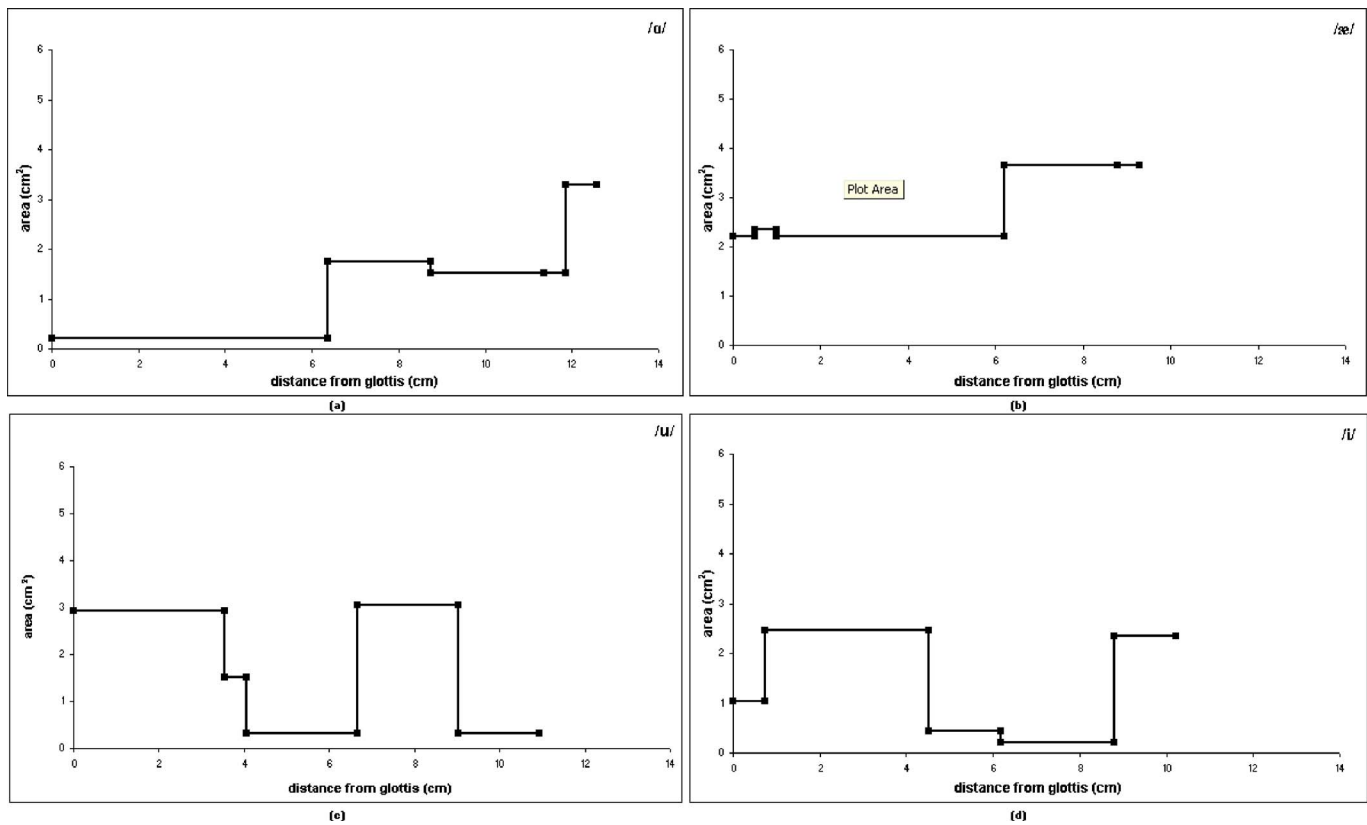
FIG. 3. Tube dimensions recovered from children's KFO condition formant frequency data (a) /ɑ/, (b) /æ/, (c) /u/, and (d) /i/.
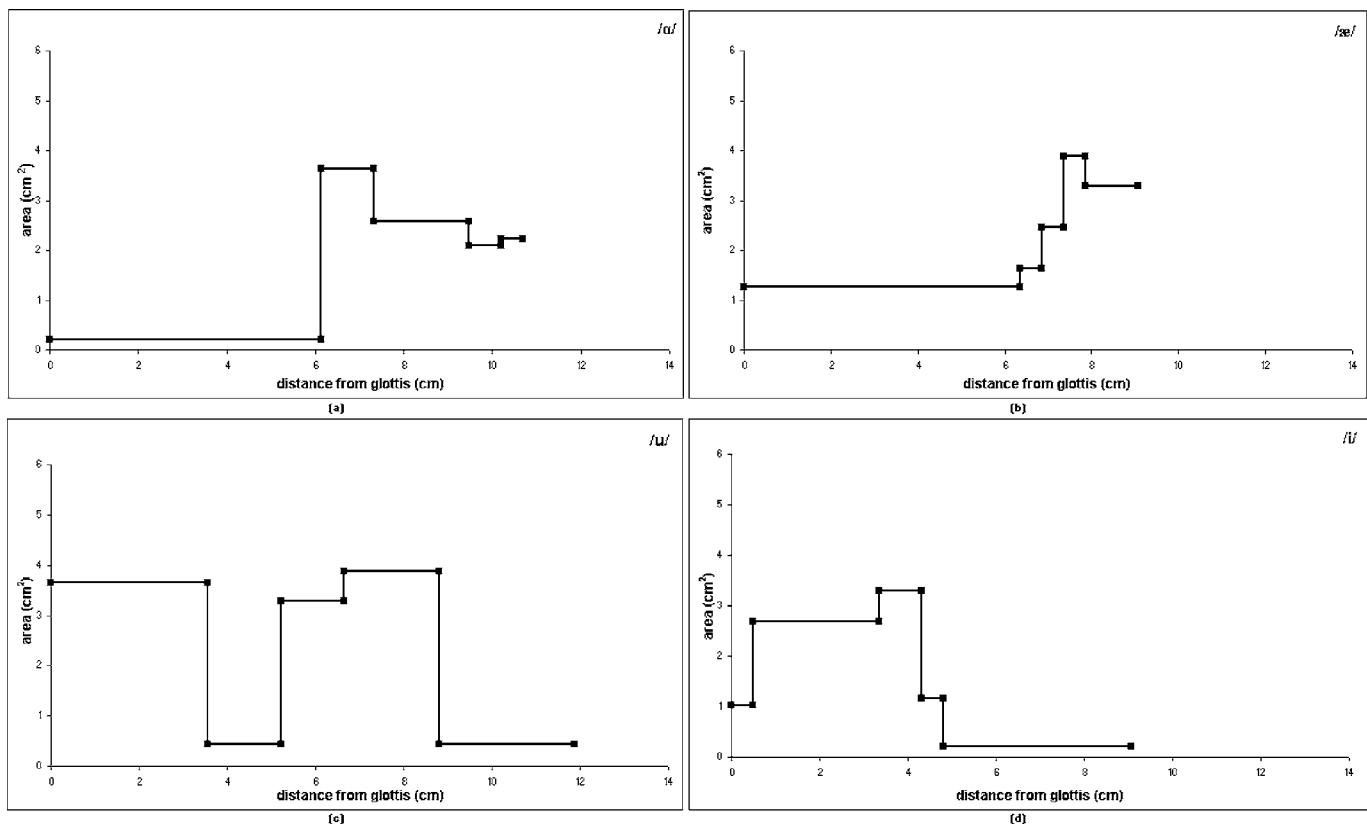
FIG. 4. Tube dimensions recovered from children's listening condition formant frequency data (a) /ɑ/, (b) /æ/, (c) /u/, and (d) /i/.

TABLE IV. The length of the recovered rear tube and its ratio with the total vocal tract length. Multiple values appear when there were multiple recovered shapes. Exceptions to when the children's rear tube-to-total length ratio is greater than or equal to the same ratio for adults are shown in italics. (When there was a tube of intermediate area between the rear tube and the third tube from the rear, one half of its length was added to the rear tube length to provide the number in this table.)

	Adult length and ratio	Children (listening) length and ratio	Children (KFO) length and ratio
ɑ	4.89 cm, 0.34	6.13 cm, 0.57	5.89 cm, 0.53
	7.09 cm, 0.43		
	6.28 cm, 0.41		
	8.75 cm, 0.51		
æ	4.75 cm, 0.31	9.06 cm, 0.78	6.19 cm, 0.67
	7.57 cm, 0.50		
u	3.08 cm, 0.17	3.55 cm, 0.30	3.80 cm, 0.35
i	6.62 cm, 0.43	9.30 cm, 0.49	4.51 cm, 0.44
	5.94 cm, 0.39		
	7.11 cm, 0.48		
	6.65 cm, 0.45		
	7.09 cm, 0.49		

constriction for /ɑ/ and an unconstricted cavity for /u/.) In the case of /u/, the difference in proportionate terms can be as large as a factor of 2, and in absolute terms, the rear tube for adults and for children is of similar length. One result of the proportionate difference in rear tube length is the possibility that the natural frequency of the rear Helmholtz resonator in children is less than the frequency of their front Helmholtz resonator, where the opposite is the case for adult's production of a "canonical" /u/.

The following can be offered as one cause for the relatively forward tongue constriction in /u/ for a 4 year old talker. If the child's constriction were moved any further back, it may be that the rear cavity would be so small or totally obliterated, that a double Helmholtz resonance would not be possible. This, in turn, may be due to the limited control and/or flexibility of the tongue body and root, and the fact that the pharynx tube forms an obtuse angle with the mouth cavity, while adult pharynges form right angles with the mouth (Kent and Vorperian, 1995, p.161). (The more obtuse the angle between the pharynx and mouth, the closer the pharynx and mouth are to forming a straight tube, and the more tongue body needs to bend to form the rear cavity and constriction.) Corroborating research on the production of [ɹ] by young children indicates that it is difficult for them to make two simultaneous constrictions with the tongue, and that the rear constriction for [ɹ] may be missing or not par-

TABLE V. Helmholtz resonance frequencies for the uncoupled back and front resonators.

	Adults resonance freq. (Hz)	Children (listening) resonance freq. (Hz)	Children (KFO) resonance freq. (Hz)
Back resonator	723	784	564
Front resonator	357	575	832

ticularly tight (McGowan, Nittrouer, and Manning, 2004). Thus, the child produces an acceptable /u/ with a proportionately longer rear cavity compared to the adult double Helmholtz resonator.

The results presented here are in accord with the finding that simple area function scaling from adult to child is not sufficient to account for the observed vowel formant frequencies. For instance, Nordström (1979) attempted to reproduce children's formant patterns by scaling the tube lengths for adults' area functions according to the ratio of children-to-adults pharyngeal length and mouth length separately. As this did not provide sufficient fit to children's vowels, the cross-sectional areas in the pharyngeal and mouth regions were, themselves, scaled by the squares of these scaling factors. This also did not provide a good fit to children's formant frequencies.

Scaling based on two anatomical dimensions is not sufficient to explain the differences in adults and children in their vowel production. Possible physiologic causes have been posited in this paper for the differences in /u/ production, coupled with an anatomic difference: flexibility of the tongue, control of the tongue body and root, and the oblique angle between the pharynx and the mouth. There are further factors to consider, including the mandible, which is disproportionately smaller for children four years of age (Kent and Vorperian, 1995). There may be other factors that prevent proportionate or encourage disproportionate formant scaling, as in the case of /u/. Further, /æ/ exhibits nonuniform formant scaling, but a physical reason has not been posited here. It is intriguing that, as Maeda and Honda (1994) point out, /u/ and /æ/ are extreme vowels for the styloglossis-genioglossis anterior antagonist pair. Further work should study /æ/ in relation to other relatively open vowels, such as /ɛ/ and /ʊ/.

The findings here do not corroborate the speculation made by Maeda and Honda (1994) that one needs the right angle of the rear pharyngeal wall and the mouth to produce the vowels /ɑ/ and /i/. (It should be born in mind, however, that the present study does not use articulatory constraints.) Table III shows that listeners judged the children's vowels under the KFO condition to be good exemplars of their phonemic categories. Even without a pharynx at right angles to the mouth, the average 4 year old child can produce these vowels in a ways that are perceptually acceptable to adults. In fact, Kent and Forner (1979) show that the vowels /ɑ/ and /i/ are produced with scale factors that are fairly constant across the three formants (Table I). The results here indicate that in order for children to achieve proportionate formant scaling it is important to form constrictions and cavities whose lengths are not far from proportionate to the adults. (A statement that proportionate length is sufficient cannot be made because of differences in the associated cross-sectional areas.) The vowels /ɑ/ and /i/ show that mismatches in the relative positions and sizes of fixed anatomical structures, such as the pharynx and mouth, do not necessarily hinder a child's ability to form such constrictions and cavities. Indeed, in perceptual tests using speech produced by an articulatory synthesizer, Ménard, Schwartz, and Boë (2004) showed that infants can produce all the French vowels. How-

ever, caution should be used in this kind of experiment, because scaling of an adult vocal tract model to simulate a child's vocal tract is not sufficient to ensure that all the important physiological aspects of children's speech are captured.

This research offers no clear strategy that a child may adopt in attempting to imitate an adult's vowels. It is plausible that the child attempts proportionate scaling of tube lengths in certain instances, and when this is impossible, he or she attempts to imitate the general vocal tract shape. It is not possible to say from these data whether children attempt to imitate vocal tract shape or acoustic parameters. This may be impossible in any study that involves only acoustic and/or articulatory data because of the causal relation between the two domains.

ACKNOWLEDGMENTS

APPENDIX

The choices given to the listener when the synthetic vowel was /ɑ/, or a vowel with its formant frequencies up to 20% different, were: "hot," "haughty," "hut," "hat," "how" or /ɑ, ɔ, ʌ, æ, oʊ/.

The choices given to the listener when the synthetic vowel was /æ/, or a vowel with its formant frequencies up to 20% different, were: "had," "head," "hot," "hayed," "hut" or /æ, ɛ, ɑ, eɪ, ʌ/.

The choices given to the listener when the synthetic vowel was /u/, or a vowel with its formant frequencies up to 20% different, were: "whoed," "hood," "hoed," "hut," "hot" or /u, ʊ, oʊ, ʌ, ɑ/.

The choices given to the listener when the synthetic vowel was /i/, or a vowel with its formant frequencies up to 20% different, were: "heed," "hid," "hayed," "head," "had" or /i, ɪ, eɪ, ɛ, æ/.

[1]Rear cavities were defined using the following rules: (1) Rear tubes are constrictions for /ɑ/ and /æ/, and they are cavities for /i/ and /u/, (2) going from rear to front, a constriction (cavity) never ends at a tube whose neighbor to the rear has a greater (lesser) cross-sectional area, (3) cross-sectional area differences of less than 0.5 cm^2 are ignored, (4) going from rear to front a constriction (cavity) ends at the first tube with increased (decreased) area, and (5) if the front neighbor of the tube ending a constriction (cavity) has an even greater (lesser) area, than one-half the length ending the constriction (cavity) is added to the length of the rear tube, otherwise there is no extra length added to the rear tube.

Eguchi, S., and Hirsh, Il J. (**1969**). "Development of speech sounds in children," Acta Oto-Laryngol., Suppl. Suppl. 257, 5–48.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G. (**1975**). "Non-uniform vowel normalization," Speech Transmission Laboratory, KTH. STL-QPSR 2-3/1975, pp. 1–19.

Fitch, W. T. and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Goldberg, D. E. (**1989**). *Genetic Algorithms* (Addison-Wesley, Reading).

Goldstein, U. (**1980**). "An articulatory model for the vocal tracts of growing children," Ph.D. dissertation, MIT, Cambridge, MA.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Kent, R. D. and Forner, L. L. (**1979**). "Developmental study of vowel formant frequencies in an imitation task," J. Acoust. Soc. Am. **65**, 208–217.

Kent, R. D. and Vorperian, H. K. (**1995**). "Development of the craniofacial-oral-laryngeal anatomy: A review," J. Medical Speech-Language Pathology **3**, 145–190.

Ladefoged, P. and Maddeison, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford).

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of spectral and temporal parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

McGowan, R. S. (**1994**). "Recovering articulatory movement from formant frequency trajectories using task dynamics and a genetic algorithm: Preliminary results," Speech Commun. **14**, 19–49.

McGowan, R. S., Nittrouer, S., and Manning, C. J. (**2004**). "Development of [ɹ] in young, Midwestern, American Children," J. Acoust. Soc. Am. **115**, 871–884.

Maeda, S., and Honda, K. (**1994**). "From EMG to formant patterns of vowels: The implications of vowel spaces," Phonetica **51**, 17–29.

Ménard, L., Schwartz, J.-L., and Boë, L.-J. (**2004**). "Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized French vowels from birth to adulthood," J. Speech Lang. Hear. Res. **47**, 1059–1080.

Nordström, P.-E. (**1979**). "Attempts to simulate female and infant vocal tracts from male area functions," Speech Transmission Laboratory, KTH. STL-QPSR 2-3/1975, pp. 20–33.

Olive, J. P., Greenwood, A., and Coleman, J. (**1993**). *Acoustics of American English Speech* (Springer-Verlag, New York).

Peterson, G. E. and Barney, H. L. (**1952**). "Control Methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, Massachusetts).

Modeling coupled aerodynamics and vocal fold dynamics using immersed boundary methods

Comer Duncan[a] and Guangnian Zhai[b]
Department of Physics and Astronomy, Bowling Green State University, Bowling Green, Ohio 43403

Ronald Scherer[c]
Department of Communication Disorders, Bowling Green State University, Bowling Green, Ohio 43403

The penalty immersed boundary (PIB) method, originally introduced by Peskin (1972) to model the function of the mammalian heart, is tested as a fluid-structure interaction model of the closely coupled dynamics of the vocal folds and aerodynamics in phonation. Two-dimensional vocal folds are simulated with material properties chosen to result in self-oscillation and volume flows in physiological frequency ranges. Properties of the glottal flow field, including vorticity, are studied in conjunction with the dynamic vocal fold motion. The results of using the PIB method to model self-oscillating vocal folds for the case of 8 cm H_2O as the transglottal pressure gradient are described. The volume flow at 8 cm H_2O, the transglottal pressure, and vortex dynamics associated with the self-oscillating model are shown. Volume flow is also given for 2, 4, and 12 cm H_2O, illustrating the robustness of the model to a range of transglottal pressures. The results indicate that the PIB method applied to modeling phonation has good potential for the study of the interdependence of aerodynamics and vocal fold motion. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2354069]

Pages: 2859–2871

I. INTRODUCTION

Computational modeling of the dynamics of phonation is a challenge from the standpoint of the inherent interdependence of glottal material properties, the motion of the vocal folds, and the complex aerodynamics that develops in the glottis. The main approaches which have been used to date may be generally classified as (1) lumped parameter models using multiple masses (Story and Titze, 1995), (2) finite element based models (Alipour, Berry, and Titze, 2000), and (3) a finite element based biophysical model (Alipour and Titze, 1996; Alipour and Scherer, 2000; Alipour and Scherer, 2004). In the first type of modeling the vocal folds are approximated by a few (two or three) lumped masses which undergo motion transverse to the inflow direction. The focus in this type of model is the dynamics of the lumped masses which model the vocal folds. The aerodynamics is modeled assuming a steady-state based approximation. The second approach has focused on the development of a more sophisticated finite element model of the vocal folds. The emphasis was on the modal properties of the finite element based material model. The aerodynamics was included in a restricted steady-state treatment. In the third approach the aerodynamics was treated using a Navier–Stokes model for the air and the material was treated using a finite element model for the vocal folds. While all three approaches have treated important parts of the system, none other than the biophysical model method of Alipour and co-workers allow the potential complexity of the model to approach the real system. The method presented in this work is a step toward a more complete model with special attention to the closely coupled interactions between air and vocal folds. The approach taken in the present work is unique in its treatment of the interface between the vocal folds and the air. This work demonstrates how emergent biophysical properties of both the air and the vocal fold can result from an intrinsic treatment of the interface.

This work focuses on simulations which utilize the penalty immersed boundary (PIB) method developed by Kim (2003), based on the immersed boundary method originally developed by Peskin (1972) and further elaborated by Peskin and McQueen (1989) for the heart system. The PIB method as applied in this work treats the dynamics of the air and the vocal folds on a more or less equal footing. The emergent complex dynamics of the system comes from the coupled nature of the two interacting subsystems, and the PIB method described in this work shows promise in providing a more intrinsic characterization of the physics of phonation.

In the work reported here, some of the basic features of the PIB method are presented, a description of a simplified two-layer treatment of a two-dimensional model of the material properties of the modeled vocal folds is given, and a summary of the results of simulations of phonation using the combined system are given. Snapshots of the vocal fold dynamics and the aerodynamics generated by the interactions among the subsystems are shown. The 8 cm H_2O case was used throughout because this transglottal pressure is common in human phonation and should result in sustained oscilla-

_____

[a]Electronic mail: comer.duncan@gmail.com
[b]Electronic mail: gnzhai@yahoo.com
[c]Electronic mail: ronalds@bgsu.edu

tions. The results indicate that this simplified two-dimensional model yields interesting interrelations between the closely coupled dynamics of the vocal folds and the aerodynamics.

II. PIB METHOD APPLIED TO MODELS OF PHONATION

The basic idea of this approach to modeling the coupled air-vocal folds system is to approximate the air as a viscous, incompressible fluid in which there is an immersed material boundary, here taken to be the vocal folds. The method handles the coupled system of air and vocal folds as a single, potentially strongly interacting whole. In general, the coupling can involve quite complex motions of both the air and the vocal folds. The dynamics that emerges from this interaction cannot be determined from approximations that assume one or the other of the subsystems dominates. As in the physiological system, in the computational model the behavior must emerge from the complex interplay between the two subsystems. This is a strength of the method that this work utilizes.

A. Summary of the PIB method

The PIB method of Kim and Peskin (Kim, 2003), when applied to modeling the vocal folds and glottal aerodynamics, contains three interrelated subsystems: (a) the air, (b) the immersed material, and (c) the interface between them. The PIB method is in the general class of methods which addresses the interacting systems (air and vocal folds) as a fluid-structure interaction problem. Thus the method must detail its treatment of the interface between the air and vocal folds, which in turn determines how the two subsystems influence each other.

1. The air

The air is assumed to satisfy the Navier–Stokes equations for a viscous, forced, incompressible fluid

$$\rho\frac{\partial u}{\partial t} + \rho u \cdot \nabla u = -\nabla p + \mu\nabla^2 u + f,$$

$$\nabla \cdot u = 0, \tag{1}$$

where ρ, $\mathbf{u}$, p, and μ are, respectively, the density, velocity, pressure, and viscosity of the air. The term $\mathbf{f}$ is the force density (force per unit area in the current two-dimensional model) used in this paper. It is the force density exerted on the air by the vocal folds at the interface between them. This force density is related to the force per unit length $\mathbf{F}$ on the vocal folds at a given point along the interface labeled by s due to internal forces in the material via

$$f(x,t) = \int F(s,t)\delta^{(2)}[x - X(s,t)]ds, \tag{2}$$

where $\delta^{(2)}$ is the two-dimensional Dirac delta function. $\mathbf{X}(s,t)$ represents the position vector of a point on the vocal fold parametrized by s at time t. This provides an important link between the two subsystems. The forced Navier–Stokes
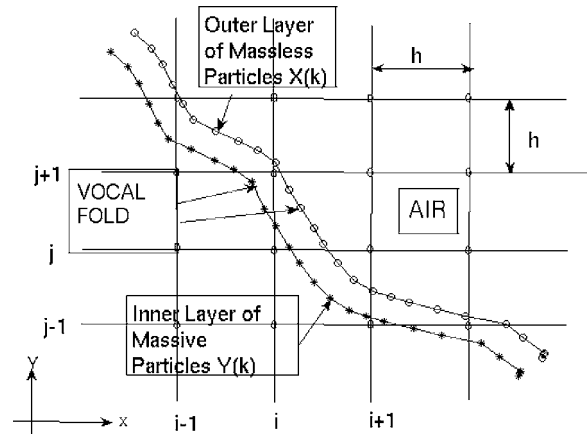
FIG. 1. Eulerian grid for aerodynamics and Lagrangian grids for vocal fold material model. The Eulerian mesh is a fixed Cartesian mesh the nodes are labeled by i, j with $i=1, nx$ and $j=1, ny$. The velocity and pressure fields are stored at the nodes labeled by i, j. The Lagrangian mesh is used to describe the moving particles which constitute the model vocal folds. On the Lagrangian mesh the particles are labeled by their Cartesian coordinates referred to the background Eulerian mesh. The positions for the outer layer of massless particles adjacent to the air are labeled by $X(k)$ with $k=1, n_{\text{particles}}$. The positions of the inner layer of particles which have mass M are labeled by $Y(k)$.

equations are set on a fixed background Eulerian mesh using Cartesian coordinates with mesh of constant spacing h in two dimensions in the present work. Figure 1 illustrates the node-centered storage for the velocity $\mathbf{u}$ and pressure p characterizing the aerodynamics.

2. The vocal fold material

The vocal fold "material" adopted in the present paper requires choosing appropriate models of the internal forces between the immersed material parts. A vocal fold is modeled by a two-layer system of particles, as is shown in Fig. 1. These two sets of particles constitute the outer and inner layers, respectively, of our simplified two-layer material model. The PIB method utilizes both Eulerian and Lagrangian meshes. The Eulerian mesh is a fixed background mesh for the description of the aerodynamic velocity and pressure fields, which are to satisfy the forced Navier–Stokes equations. The Eulerian mesh is Cartesian with a discrete set of nodes which are equally spaced in both the x and y directions. The mesh spacing is h. Eulerian mesh points are indexed using (i,j) with $i=1, nx$ and $j=1, ny$. The velocity and pressure fields are centered at the (i,j) nodes. The Lagrangian mesh consists of two layers of particles. The outer layer of mesh points immediately adjacent to the air consists of massless particles whose positions are described using a vector array $\mathbf{X}(k)$, where k indexes the particles and runs from 1 to $n_{\text{particles}}$. A second Lagrangian mesh consists of massive particles whose positions are described by a vector array $\mathbf{Y}(k)$. The massless and massive particles are adjacent and their motion is governed by the forces exerted on them by surrounding particles. In general, the Lagrangian mesh points do not coincide with the background Eulerian mesh. The force per unit length $\mathbf{F}$ on a given material point in Eq. (2) is the sum of two parts: the force per unit length on the element at $\mathbf{X}(k)$ due to adjacent elements, denoted by $\mathbf{F}_{\text{internal}}$, deter-

mined by the material properties of the vocal folds, and the force $\mathbf{F}_M$ on the massless boundary particle labeled by $\mathbf{X}(k)$ due to a spring connecting it to a particle of mass M residing at $\mathbf{Y}(k)$. The specification of $\mathbf{F}_{\text{internal}}$ and $\mathbf{F}_M$ will be given in the following section.

3. Interface between the air and the vocal fold, and tissue forces

A small portion of the material of the vocal fold is illustrated in Fig. 1 by the two sets of particles. The massless particles move with the interpolated average of the velocity field values of the surrounding air so that the no-slip condition

$$\frac{\partial X}{\partial t}(s,t) = u[X(s,t),t] = \int u(x,t)\delta^{(2)}[x - X(s,t)]dxdy \quad (3)$$

is satisfied. In Eq. (3) the two-dimensional Dirac delta function $\delta^{(2)}$ is used since the current model is two dimensional. Thus in the PIB method, the positions of the outer layer of massless particles is adapted so that they always move with the aerodynamic velocity field interpolated to the instantaneous positions of each massless particle. This provides a useful way in which the outer layer of particles maintain the same velocity as the local air velocity.

The particles of the inner layer possess mass and are moved around subject to a spring force exerted on them via connection with the adjacent massless outer layer markers. This force is modeled by a simple spring force per unit length: $F_M = -K_M(X - Y)$. The spring constant K_M is chosen large enough so that the inner layer tracks the outer layer close enough to maintain the integrity of the two layers throughout the simulation.

The motion of the particle with mass M at $\mathbf{Y}(k)$ in Fig. 1 is governed by the equation of motion

$$M\frac{\partial^2 Y}{\partial t^2} = -F_M. \quad (4)$$

Equations (3) and (4) are discretized using second-order accurate approximations when implemented in the algorithm to update the massless and massive particles in the PIB scheme.

In the present work there are forces exerted on a given outer layer particle due to surrounding outer layer particles plus a force on a given outer layer particle do to the inner layer particle. These consist of

(a) a stretching force per unit length on a particle due to nearest neighboring particles in the outer layer

$$F_{\text{stretch}}(k) = \frac{[\tau(k + 1) * \hat{e}(k + 1/2) - \tau(k) * \hat{e}(k - 1/2)]}{ds},$$

$$\tau(k) = s_{\text{stretch}}(k) * [1(k) - ds],$$

$$1(k) = |X(k) - X(k - 1)|,$$

where $\tau(k)$ is the tension at particle k, $\hat{e}(k+1/2)$ is a unit vector centered between the particles pointing from particle $k-1$ to particle k, and $1(k)$ is the distance between particles $k-1$ and k. The parameter $s_{\text{stretch}}(k)$ is the stretching force constant.

(b) a bending force per unit length

$$F_{\text{bending}}(k) = s_{\text{bending}}(k) * \{\text{curv}[X(k)] - \text{curv}[X_0(k)]\},$$

$$\text{curv}[X(k)] = \frac{[X(k - 1) - 2 * X(k) + X(k + 1)]}{ds^2},$$

where $s_{\text{bending}}(k)$ is the bending force constant, $\text{curv}[X(k)]$ is an approximation to the second derivative of the kth particle position along the boundary and $X_0(k)$ is the initial position of the kth particle.

(c) a fixed-point restorative force per unit length promoting the return of the particle position to its initial state

$$F_{\text{restore}}(k) = - s_{\text{restore}}(k) * [X(k) - Y(k)],$$

where $s_{\text{restore}}(k)$ is the force constant, $X(k)$ is the kth massless particle position, and $Y(k)$ is the position of the inner massive particle.

In addition, there is the spring force per unit length $\mathbf{F}_M(k)$ exerted on the outer layer particle due to the inner layer particle adjacent to it

$$F_M(k) = - s_M(k) * [X(k) - Y(k)],$$

with $s_M(k)$ the spring constant per unit length of the spring connecting the kth massless particle at $X(k)$ to the kth massive particle at $Y(k)$. Due to the fact that these forcing terms F_{stretch}, F_{bending}, F_{restore}, and F_M must have physical dimensions of dyn/cm, it follows that the physical dimensions of the individual forcing constants are as follows: $s_{\text{stretch}}(k)$ has dimensions of dyn/cm, $s_{\text{restore}}(k)$ and $s_M(k)$ have dimensions of dyn/cm^2, while $s_{\text{bending}}(k)$ has dimensions of dyn. The bending constant is a phenomenological model constant which represents the tendency of the vocal fold material to return of its initial configuration, which has been modeled via a dependence on the second derivative of the particle position along the vocal fold interface.

These forces per unit length are illustrated in Figs. 2 and 3. Their vector sum

$$F(k) = F_{\text{stretch}}(k) + F_{\text{bending}}(k) + F_{\text{restore}}(k) + F_M(k)$$

represents the net force per unit length on particle k. This force is spread to the Eulerian grid using the delta-function interpolation characteristic of the PIB method. The force $\mathbf{f}$ resulting from the delta-function spread is what appears on the right hand side of the Navier–Stokes equations, Eq. (1). The nonzero values of this force $\mathbf{f}$ are typically confined to regions in the Eulerian mesh which are near the outer layer. In the PIB method this force $\mathbf{f}$ is the only way that the material layers exert influence directly on the Navier–Stokes based aerodynamics. Figure 4 schematically illustrates a typical initial configuration, showing the two sets of particles which constitute the model vocal folds.

The elastic constants describing the strength of these four kinds of forces and the mass of the inner layer particles determine the material properties. These force constants have been carefully chosen so that the resulting dynamics of the model exhibit behavior approximating known physiological kinematics. In addition we apply a body force upstream over the first few rows of the Eulerian computational domain with values chosen to induce the transglottal pressure gradient of
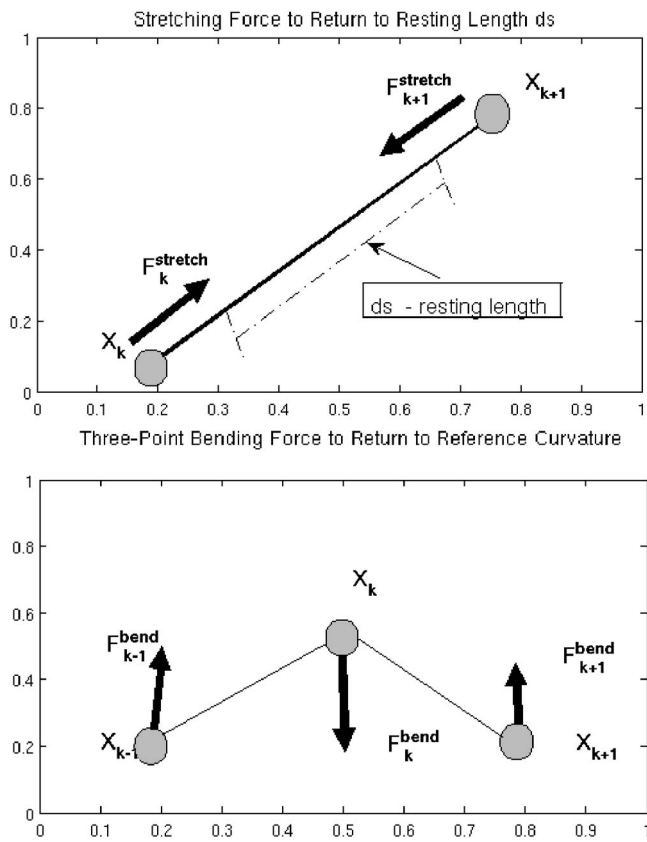
FIG. 2. (Top) Stretching force promoting return to resting length ds; (bottom) Bending force sensitive to curvature at each particle.

interest in a given simulation. In Table I the ranges of parameters chosen to characterize the material properties of the simplified two-dimensional model of the vocal fold material are shown. The values of some of the material constants vary over the extent of the model vocal folds. Figure 5 provides
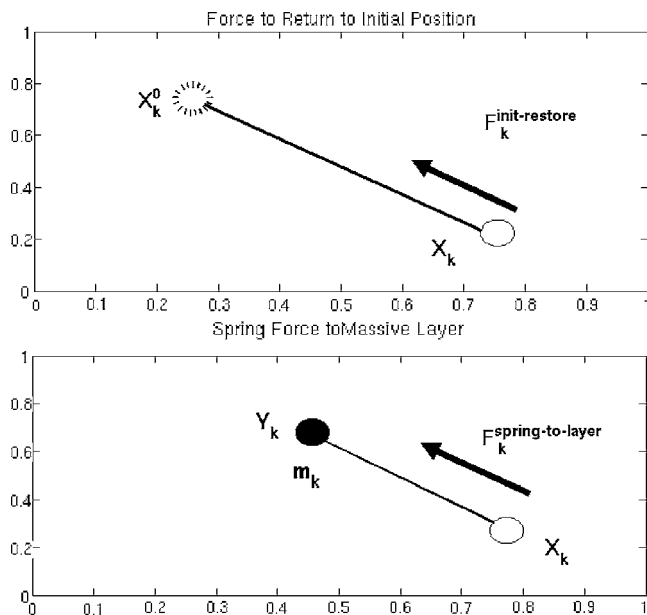
FIG. 3. (Top) Force restoring particle to initial configuration; (bottom) Force exerted by inner layer on outer layer to promote integrity of two layer mode vocal fold.
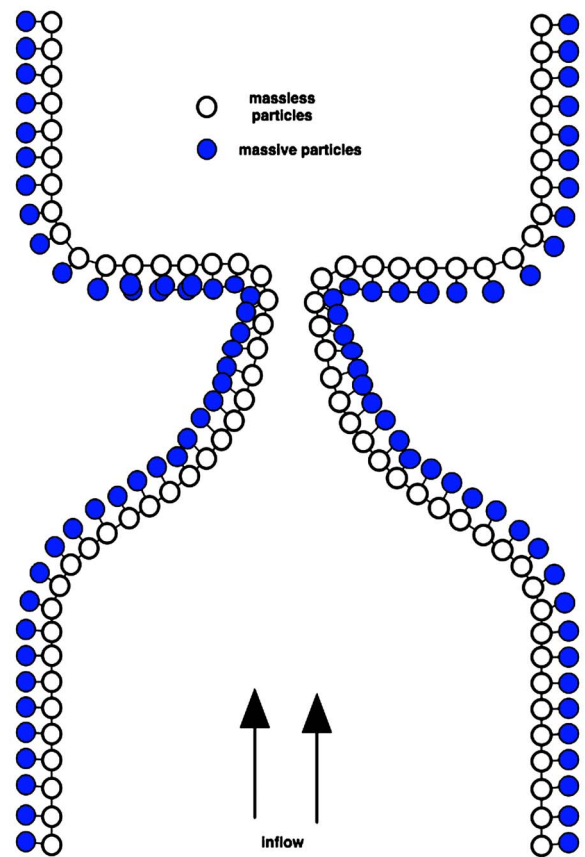
FIG. 4. (Color online) Typical initial positions of pairs of particles used to model the vocal folds material in the two-dimensional model system. Open circles represent the massless particle layer adjacent to the air. Filled circles represent the massive inner layer of particles. The inflow direction is from the bottom to the top of the diagram.

an example showing how the restoring-to-initial-configuration constant varies over the run of the initial positions of the vocal folds.

In this paper we report on a sequence of simulations with transglottal pressures of 2, 4, 8, and 12 cm H_2O, respectively. All simulations use the same material properties for the two-layer model vocal folds. The initial state of each case was chosen symmetric about the line $x=1$ cm along the y axis. All runs used the same initial adduction of 0.042 cm (minimal glottal diameter) and a parallel configuration of the left and right vocal folds. The initial velocity of the air was chosen to be everywhere zero. The total time simulated was 26.0 ms with a time step of $dt=1.0e-7$ and a sampling of the output every 0.1 ms.

All of the simulations used a fixed Cartesian Eulerian grid of 256×4096 equally spaced points in the x and y directions, respectively. The number of particles which make up the vocal folds depends on the constant Eulerian mesh spacing h in that the distance ds between adjacent particles must be no larger than $h/2$ to prevent leaking of the air

TABLE I. Material stiffness parameters.

s_M (dyn/cm^2)	s_{stretch} (dyn/cm)	s_{bend} (dyne)	s_{restore} (dyn/cm^2)	$n_{\text{particles}}$
1.0e6	1.0e6	10.0	5.0e4–1.2e6	34036
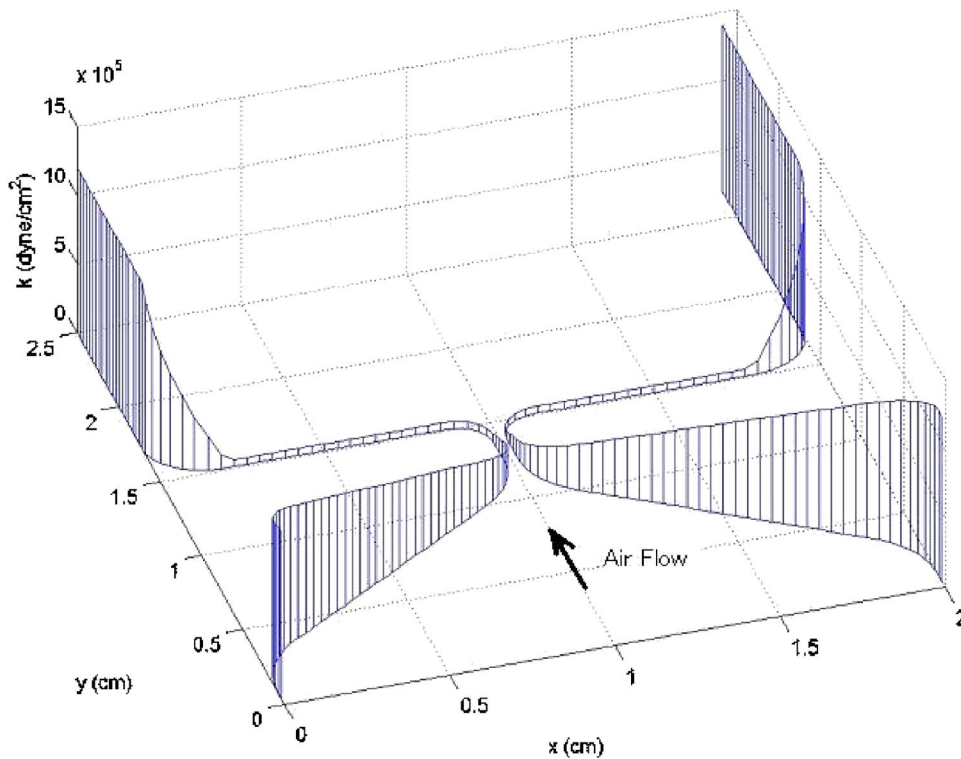
FIG. 5. (Color online) Three-dimensional plot of the restoring-to-initial-configuration parameter k of the massless boundary as a function of position for an initial configuration. The values vary over the extent of the vocal folds. The x direction is the direction transverse to the inflow and the y direction is parallel to the inflow from the bottom to the top of the diagram.

through the boundary. Typically $ds=h/4$ is chosen, as that gives some extra coverage of the vocal folds. This was useful when highly dynamic motions of the vocal folds occurred during the simulations. The Navier–Stokes solver made efficient use of the fast Fourier transform to produce velocity and pressure fields that satisfy the incompressible, forced Navier–Stokes equations. A two-step, formally second order accurate scheme described by McQueen and Peskin (2001) was used.

A single step in time from t^n to $t^{n+1}=t^n+dt$, schematically shown in Fig. 6, is outlined as follows:

(1) given the position of all boundary points and the velocity field at time t^n;
(2) update particle positions $\mathbf{X}(k)$ to time level $t^{n+1/2}$;
(3) compute the force per unit length $\mathbf{F}(\mathbf{k})$ on each vocal fold point k due to adjacent vocal fold points given by $\mathbf{F}_{\text{internal}}[\mathbf{X}(k)]+\mathbf{F}_M(k)$;
(4) spread this force to the Eulerian grid giving the force density exerted on the fluid using Eq. (2);
(5) update the Eulerian fluid velocity and pressure to time $t^{n+1/2}$;
(6) use the no-slip condition Eq. (3) to advance all the vocal fold point positions to time t^{n+1};
(7) use Eq. (1) to solve the forced Navier–Stokes equations to find the velocity and pressure field at time $t^{n+1}=t^n+dt$ and use Eq. (4) to advance the massive points at $\mathbf{Y}(k)$ to t^{n+1}; and
(8) use Eq. (4) to advance the massive points at $\mathbf{Y}(k)$ to t^{n+1}.

The simulations reported in the present work all utilized a background Eulerian mesh of 256×4096 cells, as mentioned above, and used a Lagrangian mesh of 34 036 particles with a fixed time step of $dt=1.0e\text{-}7$ s. One of the

unique features of the PIB method in the present vocal folds modeling is the high resolution that the method provides. This means that very small scale coverage of the dynamics of
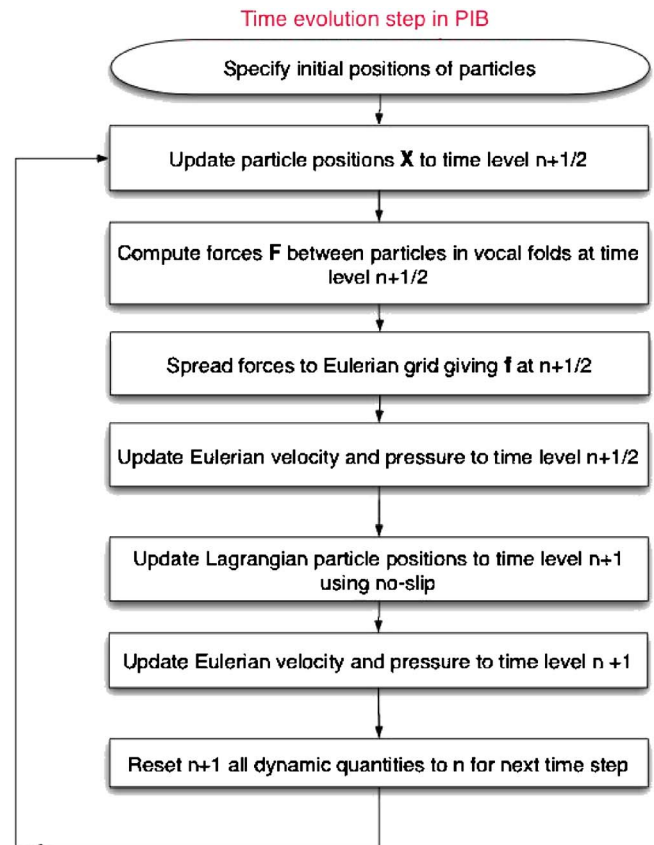
FIG. 6. (Color online) Flow chart of a single time step in the PIB method.

the vocal folds is part and parcel of the method. Thus the PIB model provides unprecedented ability to track vocal fold dynamics. The use of 256 cells in the direction transverse to the flow covering a distance of 2 cm provided adequate spatial resolution. The need for a time step of 0.1 μs was induced by the demand of numerical stability of the simulations. Thus the requirements of stability and adequate spatial coverage necessitated rather long run times, typically of the order of one week on a Pentium xeon processor.

III. RESULTS AND DISCUSSION

A. Simulation at 8 cm H$_2$O transglottal pressure

One of the important aspects of any model of phonation is its ability to exhibit self-oscillation under the condition of an appropriately chosen transglottal pressure. To illustrate some of the characteristics which come out of the PIB model, a simulation for a transglottal pressure of 8 cm H$_2$O was performed. The method uses periodic boundary conditions in both the transglottal (y) and the transverse (x) directions, as required by the Navier–Stokes solver used here. In order to induce a pressure gradient across the glottis, a body force of constant magnitude was applied over two rows of the domain near the inlet (near $y=0$). The body force magnitude was chosen so that the resulting transglottal pressure was equivalent to 8 cm H$_2$O. The initial minimal separation between the vocal folds was set to 0.042 cm and the medial vocal fold surfaces were placed in a parallel arrangement at $t=0$.

The simulation gives the positions of the particles representing the vocal folds and the velocity and pressure fields of the air. Given the substantial memory requirements for the simulation, these quantities were written out every 0.1 ms during the simulation. Space limitations of the tools for data analysis (MATLAB) made the spatial and temporal resolution coarser than the resolution simulated. In the presentation here, the spatial data sets have been down-sampled by a factor of 4 in each direction. While the resulting spatial resolution is relatively coarse, the essential aspects of the model are preserved.

1. Volume flow and transglottal pressure dynamics

The properties of the volume flow are important output of the vocal fold dynamics. Given the restriction to two dimensions in the present work, the definition used for the volume flow is given. The volume flow U_0 used in the analysis of the simulation output is defined in general by

$$U_0 = \int \frac{dV}{dt},$$

where dV is the volume swept out in time dt and the integral is over an area perpendicular to the y component of the aerodynamic velocity field. In the present case

$$\frac{dV}{dt} = u_y(x, y_{\text{fixed}}, t^n) * dx * dz.$$
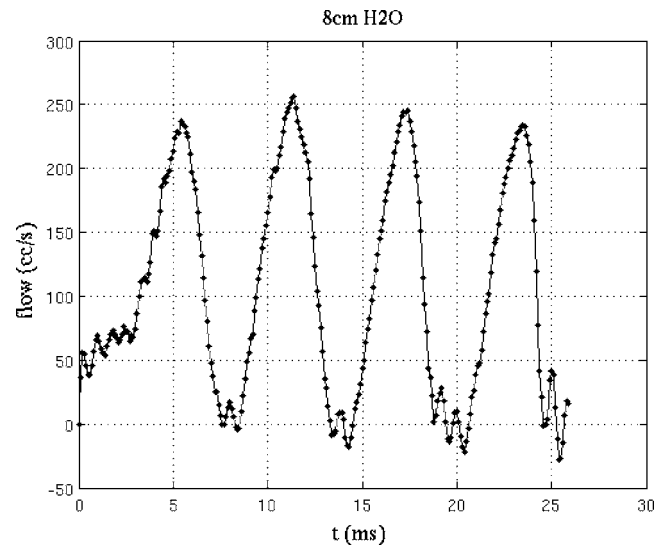
FIG. 7. Volume flow rate vs time for transglottal pressure of 8 cm H$_2$O. The dots represent the values of the volume flow rate at intervals of 0.1 ms.

The integral is over the x direction from one side of the system to the other, covering the range 0, $1x$ in x. The fixed value of y is taken to be well outside the region where the vocal fold dynamics is significant. Here it is taken to be at the y location corresponding to $j=308$, or $y=2.412$ cm. dz is taken to have a unit value. Thus the volume flow rate used in the present work is

$$U_0 = \int_0^{1_x} u_y(x, 308 * h) * dx * dz.$$

The discrete version of the integral used in the present work is

$$U_0 = dz * (dx) * \sum_{i=1}^{nx/4} u_y(i, j_{\text{fixed}}),$$

where dx has been taken to be $4*h$ since the velocity data has been sampled at every fourth spatial point in each direction, as alluded to above. The sum over i from 1 to 64 covers the range in x from 0 to $1_x=2$ cm and $j_{\text{fixed}}=77$ in the sampled data set runs.

Figure 7 presents the volume flow. Given that the data were written out in time increments of 0.1 ms, the volume flow graph contains a total of 260 snapshots over the 26 ms shown. At each snapshot the volume flow was computed along a plane at $y=2.412$ cm downstream from the inlet (0.912 cm from the glottal exit; see Fig. 5).

From the volume flow shown in Fig. 7 it is clear that there is a transitory initial interval of time as the initially quiescent vocal folds begin to move in response to the external pressure gradient. After that the flow exhibits more consistent cyclic behavior. The flow indicates that there is self-oscillatory behavior for this case. There are regions where the vocal folds are nearly closed during which the flow oscillates and goes slightly negative. This behavior is most likely due to a slight backflow introduced near closure. Further investigation is underway to study the subcycle details

Duncan *et al.*: Penalty immersed boundary model of phonation
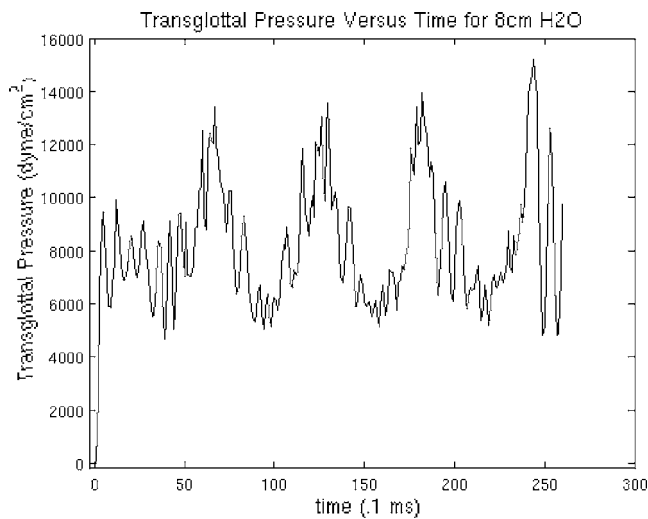
FIG. 8. Plot of transglottal pressure vs time for 8 cm H_2O.

of the closure process in the present model. It may be advantageous to introduce a collision model in the intervals during near closure.

Much dynamic behavior is going on during the complex interactions between the air and the vocal folds. A sense of the dynamics of the air may be had by observing the transglottal pressure. In the present simulations, the transglottal pressure is the difference between the pressure at $x=1$ cm, $y=0$ (center of the inlet) and $x=1$ cm, $y=2.412$ cm (see Fig. 5). This downstream point was chosen to be representative of a distance at which the flow would be clearly outside of the glottis. Since most of the pressure gradient occurs over this distance, such a difference between the two pressures should give a reasonable estimate of the effective transglottal pressure, and is expected to be a complex function of time. Indeed this is the case, as is shown in Fig. 8. An important aspect of the present PIB method as applied to vocal fold and aerodynamic modeling is that the output dynamics is determined by the detailed interaction and mutual influence of the vocal fold and the air. Such an intrinsic coupling means that the emergent dynamics is a truly cooperative phenomenon. This unique aspect of the PIB method means that it is possible for these closely coupled subsystems to induce dynamics on small spatial and temporal scales unachieved by other methods. The use of very small spatial separations between the particles (of the order of $h/4$, where h is the background Eulerian mesh spacing) provides the PIB model with a capability to model to track tissue response in greater detail than other models.

The complex dynamics of the transglottal pressure show quasi-periodic behavior with a dominant ripple frequency. The first 5 ms exhibits a complex, organized variation of the transglottal pressure that corresponds directly to the volume velocity variations of Fig. 7. Some of this behavior is associated with the initial transitory interval (covering up to 3.2 ms from the perspective of Fig. 7) with the rest continuing until the first indication of a transglottal pressure peak is seen at around 6.7 ms. The transitory interval is the result of the starting conditions and the necessary time interval through which the two-way interaction between the external

transglottal pressure applied and the biomechanical properties of the vocal fold material come to a dynamically consistent compromise. From then on, the behavior is still quite complex, but repetitions of peaking behavior reappear. The high pressure is in the range of 14 000–15 200 dyn/cm^2 (14.3–15.5 cm H_2O). The lowest transglottal pressures dip to around 4700 dyn/cm^2 (4.8 cm H_2O), and the computed mean of the data is 8211 dyn/cm^2 (8.3 cm H_2O).

Several quantitative measures of the combined vocal fold and aerodynamics will now be presented. Figure 9 shows plots of the transglottal pressure (dp), the subglottal pressure ($psubgl$), the supraglottal pressure ($psupragl$), the volume flow (U_0), and the time derivative of the volume flow (dU/dt) for the 8 cm H_2O case.

The combined plot in Fig. 9 demonstrates several correlated aspects of the 8 cm H_2O case, all of which are consistent with real phonation. Considering that the transglottal pressure is the difference between the subglottal and supraglottal pressures, it is of interest to examine these two components separately. Note that when the subglottal pressure has a peaking indication, the supraglottal pressure tends to show a trough. This is consistent with other models and measures of these pressures (e.g., Cranen and Boves, 1985), and is due to the closing glottis generating air pressure increase below the glottis and decrease above relative to reactance of the airways. Also, the numerical value of the subglottal pressure is larger than the supraglottal pressure value. The subglottal pressure has a mean value 7648 dyn/cm^2 (7.8 cm H_2O) whereas the supraglottal pressure has a mean of −563 dyn/cm^2 (−0.6 cm H_2O), with some fluctuations. Finally, the flow derivative achieves its largest absolute values on the down-slope of the flow graph, indicating that inertive loading and vocal fold motion may be contributing to flow skewing (which can be seen in the U_g trace) as expected in human phonation.

It is of interest to compare the time variation of the transglottal pressure with the volume flow in isolation. The resulting plot of both is shown in Fig. 10, where each time series is normalized to its maximum. The peaking of the dynamic transglottal pressure lags the peaking of the volume flow. This lag is to be expected, because the peak pressures are related to the closing of the vocal folds. However, the amount of lag is not constant from period to period. The differences between the peak values of the transglottal pressure and volume flow for the data displayed in Fig. 10 are 1.14, 1.59, 0.91, and 0.796 for each of the four cycles consecutively. The relation between these two quantities and other air flow properties is thus of interest. One important quantity which helps characterize the air flow field throughout the domain outside the vocal folds is the vorticity ω, where

$$\omega = \nabla \times u.$$

In the present two-dimensional model, the only nonzero component of the vorticity is the z component. In fact, our main use of the vorticity is to track its magnitude at various points in the computational domain. To this end, snapshots of contours of the vorticity have been superimposed on the po-
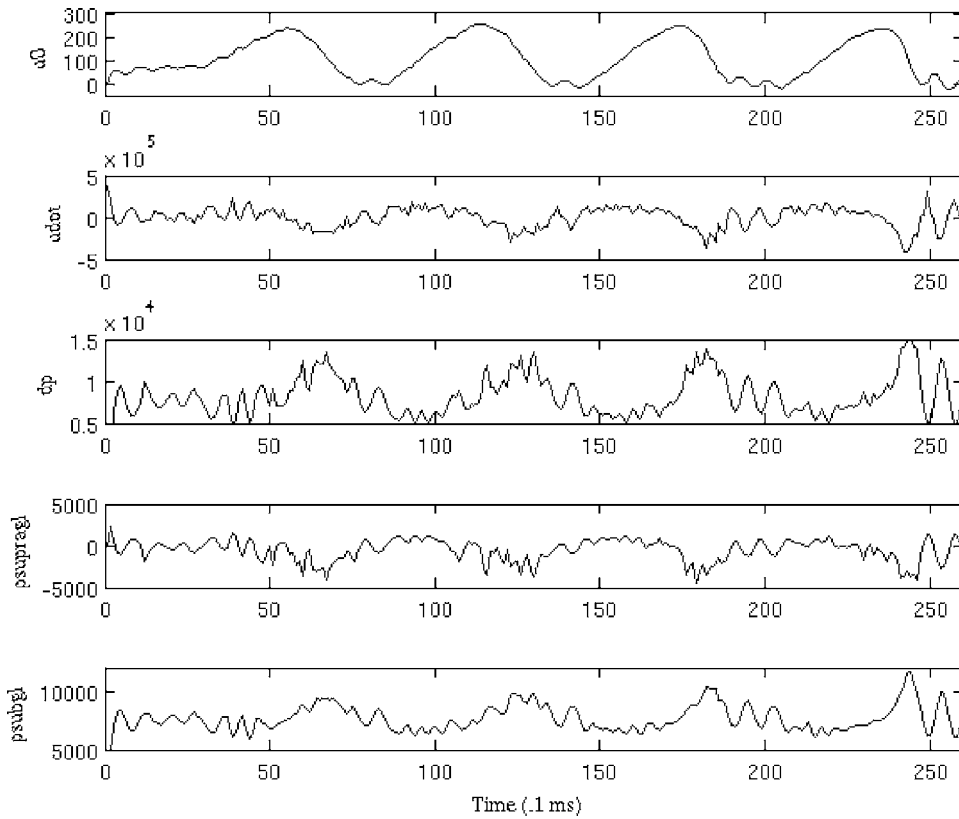
FIG. 9. Combined plots of (top to bottom) flow rate U (cm^3/s), time derivative of flow rate dU/dt (cm^3/s^2), transglottal pressure (dyn/cm^2) supraglottal pressure (dyn/cm^2), and subglottal pressure (dyn/cm^2) for an applied body force equivalent to 8 cm H$_2$O. The arrows in the U_0 graph indicate its maxima and the arrows in the transglottal pressure graph indicate its maxima.

sitions of the vocal folds in Figs. 11–14, at times corresponding to the maxima of the transglottal pressure. We note that the qualitative appearance of the glottal angle (all slightly divergent, with greater divergence with each consecutive cycle), vocal fold positions, and vorticity contours generally indicates that the decreasing flow rate is associated with a somewhat increased magnitude of vorticity. This is not to say that vorticity is not produced at other times during these cycles but that the data show a strong coincident relation
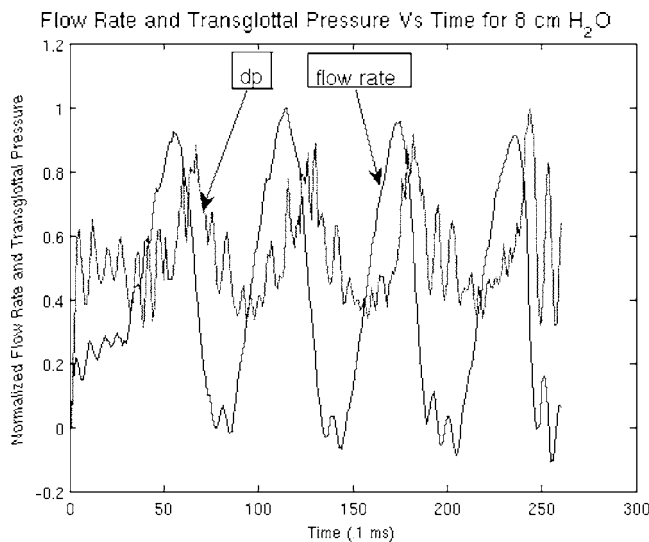
FIG. 10. Comparison of the time variation of the volume flow rate and transglottal pressure for 8 cm H$_2$O. The volume flow rate and the transglottal pressure have been normalized to their maximum values prior to plotting. The volume flow rate is shown in black and the transglottal pressure is shown in gray.
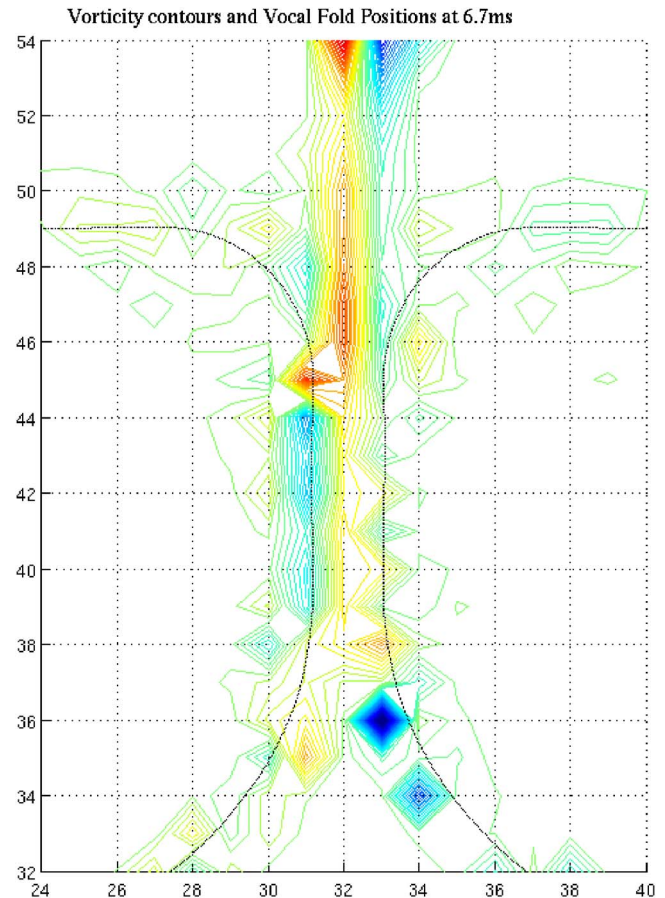
FIG. 11. (Color online) Vorticity contours and vocal folds positions at t =6.7 ms. Zoomed to cover central part of glottis. Horizontal and vertical indices represent Eulerian grid indices down-sampled by a factor of 4 in each direction.
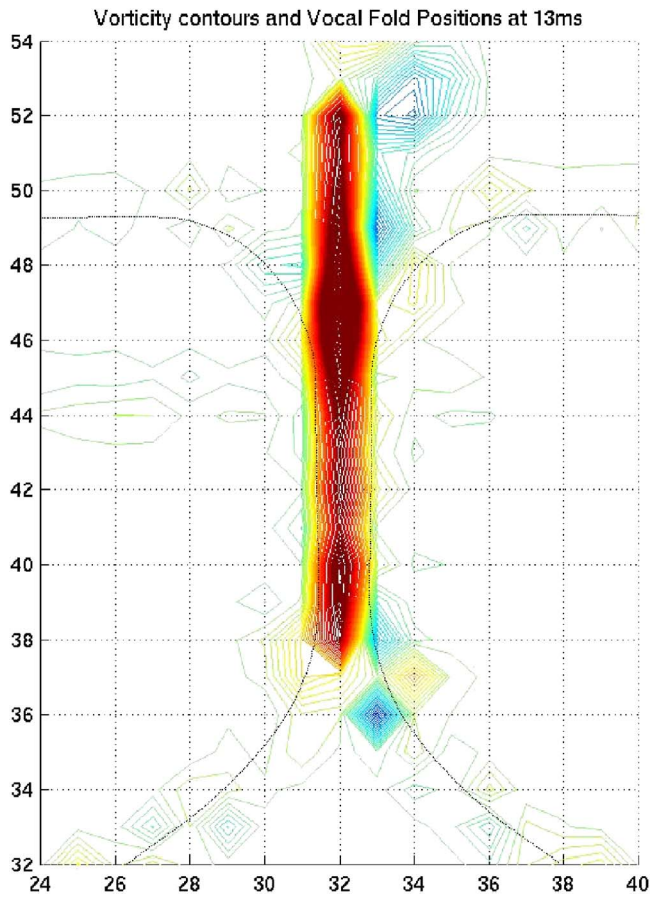
FIG. 12. (Color online) Vorticity contours and vocal folds positions at t =13.0 ms. Zoomed to cover central part of glottis. Horizontal and vertical indices represent Eulerian grid indices down-sampled by a factor of 4 in each direction.

FIG. 13. (Color online) Vorticity contours and vocal folds positions at 18.2 ms. Zoomed to cover central part of glottis. Horizontal and vertical indices represent Eulerian grid indices down-sampled by a factor of 4 in each direction.

among three aspects: transglottal pressure achieving maxima, largest negative slope (maximum flow declination rate) during the fall of the volume flow, and times of increasing magnitude of vorticity. The dynamic relation between the measures of the vocal fold positions and the aerodynamics indicates that the PIB model produces a complex dynamic system.

2. Study of a single period of volume flow

While the discussion above has focused on the volume flow, transglottal pressure, and vorticity in a preliminary way, vocal fold motion and vorticity for a single cycle of the volume flow will now be considered. The intent is to illustrate in snapshots the vocal fold positions and vorticity contours during the cycle between 13.6 and 18.9 ms. These are shown in Fig. 15. The seven arrows in Fig. 15 correspond, respectively, to the times 14, 15, 16, 17, 17.5, 18, and 18.5 ms during the cycle. At these times the instantaneous vocal fold positions and the contours of the vorticity in the glottal region are shown. As the volume flow rises from near zero to its maximum, the vorticity associated with these vocal fold positions and shapes also changes. At the time of maximum volume flow the vocal folds are near their maximum opening, and the glottal angle is slightly divergent. Also, these figures show the relation between vorticity pro-

duction and the details of the intra-glottal shape. In what follows, the salient features of the selected times shown in Fig. 15 are:

14.0 ms:

- vocal folds are nearly parallel and near closure,
- vorticity production is minimal,
- volume flow is low,
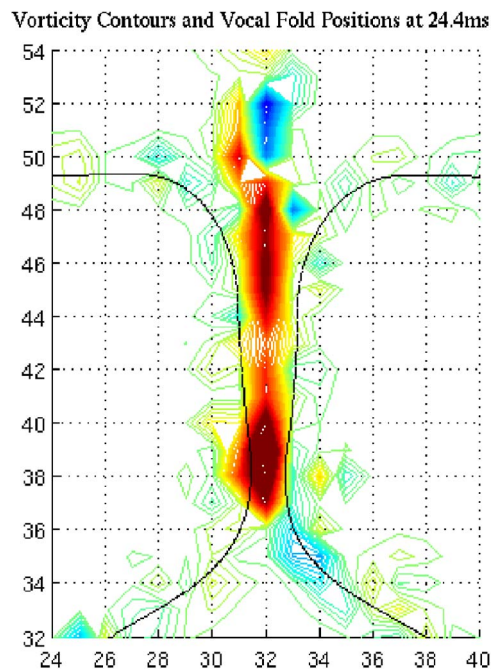
FIG. 14. (Color online) Vorticity contours and vocal folds positions at t =24.4 ms. Zoomed to cover central part of glottis. Horizontal and vertical indices represent Eulerian grid indices down-sampled by a factor of 4 in each direction.
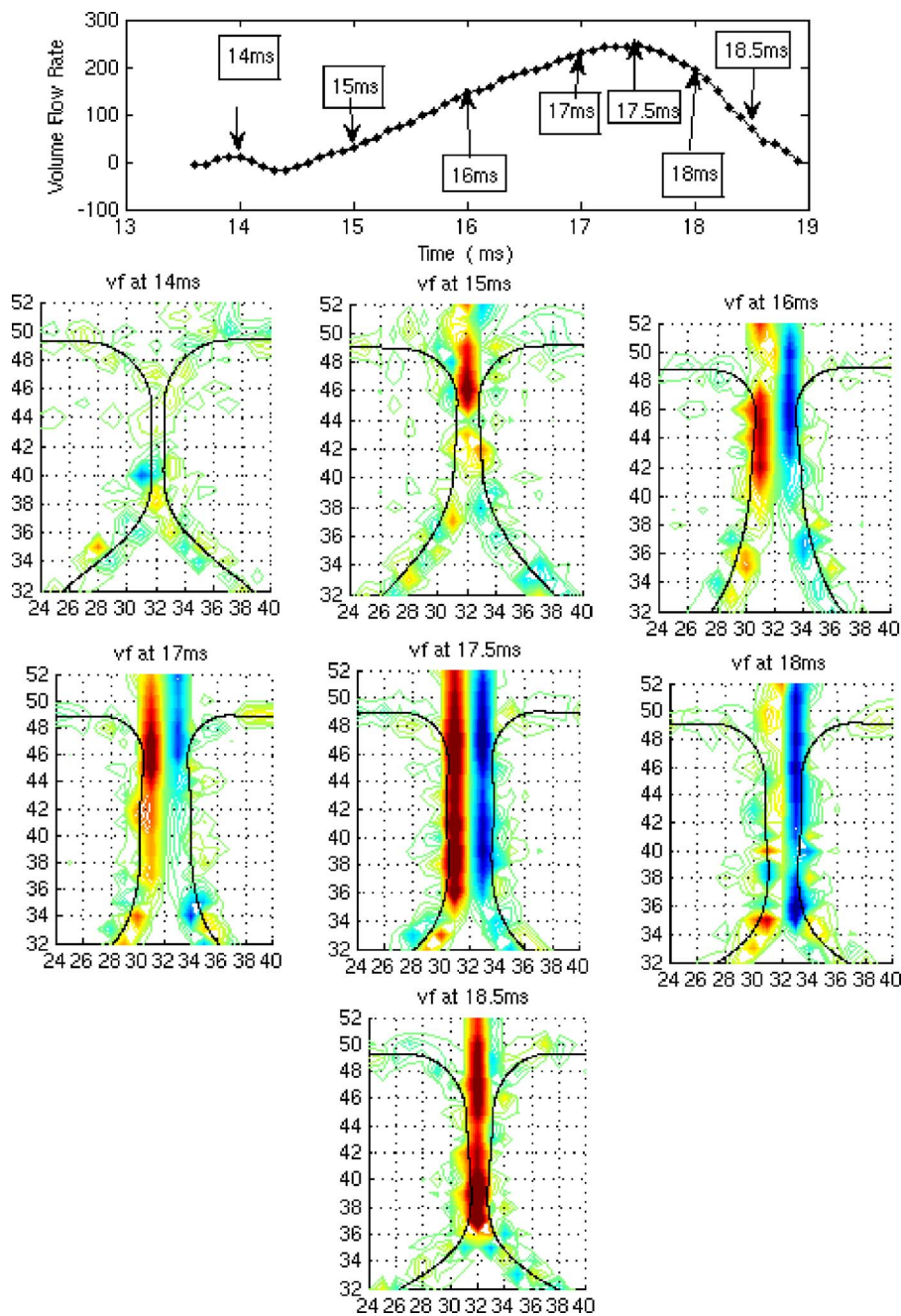
FIG. 15. (Color online) Single cycle of volume flow rate and selected vocal folds position with instantaneous vorticity contours. At the top of the figure the arrows in the volume flow correspond to times at which the overlay of the vocal fold positions and the vorticity contours: 14, 15, 16, 17, 17.5, 18, and 18.5 ms, respectively.

15.0 ms:

- vocal folds being pushed apart from below by subglottal pressure, with a convergent glottal angle and a narrower opening at the top,
- vorticity production minimal throughout glottis except at glottal outlet where the curvature of the vocal folds is significant, and
- volume flow increasing;

16.0 ms

- vocal folds are increasingly separated, glottis is convergent, and the overall diameter of the opening of the glottis is larger,
- vorticity production is associated with the outlet of the glottis, where the curvature of the vocal folds is significant, and

- volume flow increasing;

17.0 ms:

- vocal folds are near maximum separation, glottal angle is less convergent, with a notable intra-glottal curvature (concavity), indicating intraglottal shape dynamics in progress,
- vorticity production now dominant at the glottal outlet where curvature of glottis is largest, and
- volume flow increasing, as average opening of glottis is larger, but nearing the maximum value;

17.5 ms:

- vocal folds at the glottal outlet have not altered position to changed diameter and curvature substantially, but the inlet of the glottis has larger curvature; glottal angle is slightly divergent, suggesting initial closing phase,

- vorticity production associated strongly with both glottal inlet and glottal outlet regions of large curvature of glottis, and
- volume flow has maximum value;

18.0 ms:

- glottal outlet has not substantially decreased in diameter; glottis more divergent in shape,
- vorticity production associated with curvature at the glottal inlet, as curvature of that region is enhanced at this time, and
- volume flow is decreased since average glottal opening diameter is smaller;

18.5 ms:

- vocal folds moving toward each other near glottal outlet but still relatively open at the glottal outlet region; glottis more divergent with narrow glottal entrance,
- vorticity production associated with curvature of glottal inlet, and
- volume flow is decreased further.

Several general properties of the snapshots can be gleaned:

(1) as the cyle begins, the lower portion of the glottis opens first;

(2) vorticity production is associated with regions of increased curvature (i.e., smaller radii of curvature) and vorticity is largest near the top end of the glottis during the first half of the cycle; and

(3) the dynamic character of the glottal duct makes the relation between a specific point of the glottis and vorticity not an appropriate object of attention, as the glottis is changing so rapidly that the dynamic effects of the glottal wall motion on the aerodynamics are not separable. Further, given that the shape of the glottis is time dependent, it is clear that when vortical motion is in part induced by the curvature of the glottis, this curvature itself is a dynamic quantity. Thus, subsequent to the instantaneous curvature being sufficient to encourage vorticity to be produced locally, the vorticity does not find itself in a static environment. This means that if vorticity is produced, the dynamic nature of the glottal walls affects and is affected by the aerodynamics. In addition, it is of interest to estimate the skewing quotient $\mathbf{Q}_S$ and the open quotient $\mathbf{Q}_O$ (Titze, 1994; Titze and Sundberg, 1992) for the selected single cycle in the simulation. These estimates were made by first obtaining the time $\mathbf{T}_p$ during which the slope of the time derivative of the flow, dU/dt, is positive, the time T_n during which the slope of dU/dt is negative, and the time $\mathbf{T}$ for the single period. From the graph of dU/dt (Fig. 9), $\mathbf{T}_p = 3.0$ ms, $\mathbf{T}_n = 2.3$ ms, and $\mathbf{T} = 5.6$ ms. Based on the present simulation of 8 cmH$_2$O, these data provide the following estimates of the skewing and open quotients: $\mathbf{Q}_S = 1.26$, $\mathbf{Q}_O = 0.95$, values which are indicative of the initial adduction for the present simulations. Also, the value of the time $\mathbf{T}$ for the period yields a rough estimate of a frequency of 178 Hz.

An estimate was made of the maximum subglottal
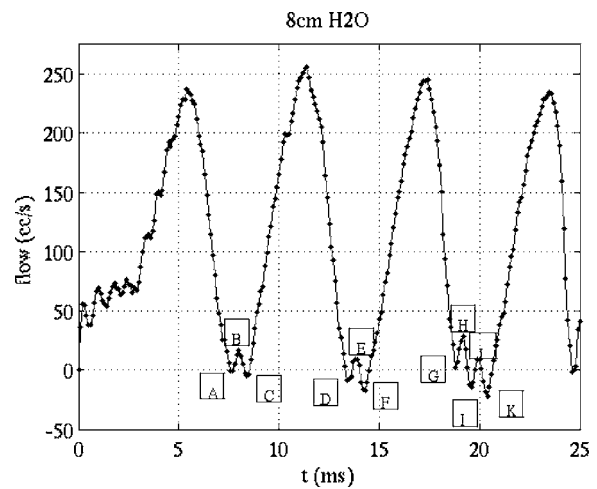
FIG. 16. Volume flow rate showing selected instants for analysis, labeled A–K. The dots represent the values of the volume flow rate at intervals of 0.1 ms.

power for the subglottal pressure of 8 cm H$_2$O, which is 784 N/m^2, and the maximum volume flow rate of 250 cm^3/s, viz.,

$$\text{Max}(\text{Power}_{\text{subglottal}}) = P_{\text{subglottal}} * U_{0\text{max}} = 0.196 \text{ W}.$$

3. Properties of the aerodynamics and vocal fold motion in regions of near zero volume flow

The volume flow graph in Fig. 7 indicates four time intervals in which the volume flow is near zero. Given the similarity of the short period oscillations in the volume flow during these intervals, the motion of the vocal folds and the attendant vorticity contours during three of these intervals were inspected. One question was, are there similarities in the aerodynamics and closely associated vocal fold dynamics that are common to all three of the intervals shown in Fig. 16?

Three separate sets of sample times, chosen by the local extrema of the volume flow where it approaches zero, are labeled (A, B, C), (D, E, F), and (G, H, I, J, K), respectively, in Fig. 16. A study of the vocal fold positions and the vorticity contours for these sets of points indicates that there is little vorticity associated with them. During these intervals, the vocal folds (and therefore the glottal walls) oscillate with rather small amplitudes, as movies of the dynamics clearly indicate. These small oscillations appear to be the cause of the local flow variations near zero, rather than, for example, resonance effects of the airway. In the simulations, the initially chosen adduction was not expected to permit complete closure. Perhaps some adjustments in one or more of the material constants would affect these "near-closing" oscillations. If the vocal fold material was a bit stiffer in one of the material constants, then the material would not be so flexible as to allow substantial suboscillations as are seen here. This possibility is under investigation and will be reported on in subsequent work.
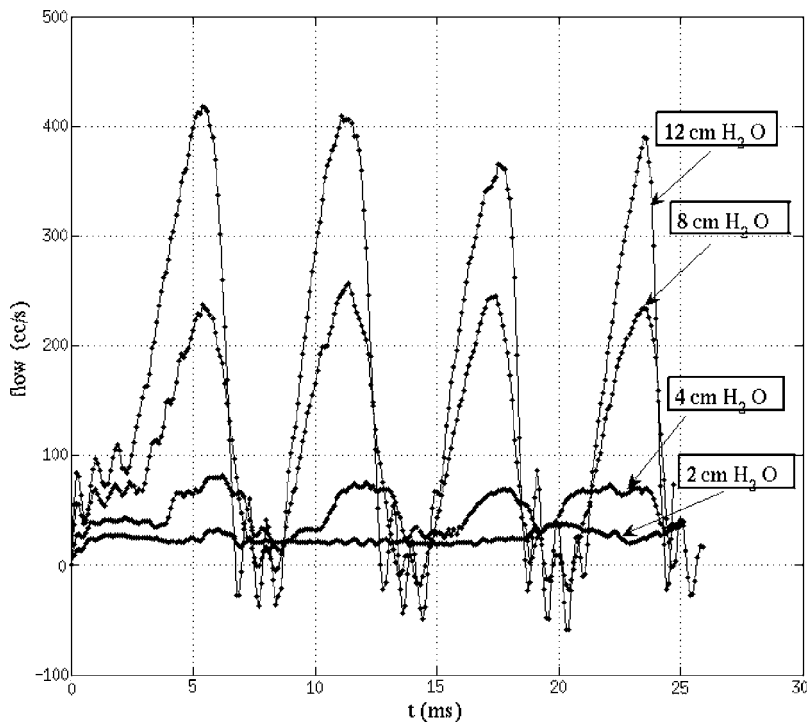
FIG. 17. Volume flow rate for 2, 4, 8, and 12 cm H_2O. The dots represent the values of the volume flow rate for the separate transglottal pressures at intervals of 0.1 ms.

B. Volume flow for several transglottal pressures

In this section the ability of the PIB model to characterize the effect of varying the transglottal pressure is explored. The main interest is to investigate the sensitivity of the simplified two-dimensional model to the onset of self-oscillation. A second reason is to inquire about how the vocal fold position dynamics and the aerodynamics associated with each different input pressure are qualitatively and quantitatively different below and above a numerically determined threshold. Such information is important to the understanding of the interrelations between the two subsystems.

Transglottal pressure values of 2, 4, and 12 cm H_2O were chosen to study the ability of the PIB model to simulate phonation and suprathreshold behavior. For the different cases, all parameters were fixed as in the 8 cm H_2O case except for the magnitude of the body force to induce the desired pressure gradient in the y direction. Figure 17 shows the volume flow graph for the four runs.

It is seen in Fig. 17 that there is a lack of significant flow modulation for the 2 cm H_2O condition, but obvious modulations for the 4 cm H_2O condition. While an apparent onset of oscillation is observed for the 4 cm H_2O case, it is presently not possible to conclude that the threshold is just above 2 cm H_2O, because intermediate pressure values were not studied, and other adduction and tissue properties should be explored. The point of the present trend in the response of the flow rate to different transglottal pressures is that the PIB model appears robust enough to study threshold behavior realistically.

IV. CONCLUSIONS AND FURTHER WORK

The results of applying the penalty immersed boundary method to a simplified two-dimensional model of the aerodynamics and vocal fold dynamics show that the method can produce physiologically relevant properties. The method allows for an intrinsic treatment of the complexities of the air-vocal fold systems and demonstrates the close connection between the production of the air flow field by the dynamic vocal fold motion. The two subsystems are not separable in the real glottis. The PIB method recognizes this inseparability and adapts to it. This preliminary study of the potential for the PIB method suggests that it should be extended to a broader range of normal phonatory conditions, and eventually to pathologies and other asymmetries of tissue biomechanics, conditions for which the model appears to be well suited. Indeed a long range goal of such modeling is to develop a comprehensive model wherein the intimate connection between the aerodynamics and the vocal fold contour and material properties will likely yield new insights into the relation between glottal structure, its dynamics, and the "output" aerodynamics. One of the challenges of modeling the vocal folds using a fixed Eulerian mesh for the velocity and pressure fields, with the Lagrangian mesh for the vocal fold material moving on that mesh, is to capture all the important physics in the model. Simplified lumped models necessarily leave out the expression of potentially physiologically important effects. Thus the PIB method initially explored in the present work frees the interacting subsystems to produce vocal fold oscillations which potentially more closely represent the real world vocal fold oscillations. It must be mentioned, however, that the present PIB model does not take into account the effects of coupling to the trachea and the pharynx (except to the extent shown in Fig. 6). Further, the influence of the false vocal folds is not included in the present study. Only after coupling to these other important subsystems can the PIB method be said to represent a complete model. Such developments are currently under active study.

As the Reynolds number associated with the flow of air increases, the number of Eulerian mesh points needed in-

creases. Even in two spatial dimensions it thus becomes clear that at some point the computational resources will be severely challenged to provide the needed spatiotemporal resolution. In order to address this problem, a program is under development to implement the immersed boundary methods in the context of parallel adaptive mesh refinement (AMR) technology coupled with a more general Navier–Stokes solver which can accommodate a variety of boundary conditions. Currently under development is a version of the two-dimensional PIB method which uses the infrastructure of methods developed at Lawrence Berkeley Labs (LBL) (Almgren *et al.*, 1998). The LBL computational technology has been developed in part to allow the implementation of a robust class of solvers of the forced Navier–Stokes equations for variable density systems. Underlying the solver is a set of meshes with varying degrees of resolution, with the meshes created as the physics being expressed requires more resolution. The meshes are thus dynamically created and destroyed as the simulation proceeds, giving needed coverage of small scale phenomena while doing so consistent with phenomena at coarser scales. The net result of this approach will be to provide more accurate characterization of the vocal fold-air system and do so in a computationally efficient manner. In addition, this technology has been parallelized. This means that the evolution of the vocal fold system will be accomplished with much enhanced spatial and temporal resolution and carried out on large parallel clusters. This has an important practical payoff: by spreading the computations among a large number of processors, more resolution is obtained than with a simulation bound to just one processor. Such a coupling of parallel AMR with a robust class of Navier–Stokes solvers is expected to yield increased resolution and computational efficiency.

The present PIB model has another simplification which would be removed in subsequent studies. This is that the material model has no viscous tissue forces. While physiological vocal fold tissue possesses viscosity, it was felt that for the initial study the effect of such dissipation would not be included. By not including viscosity of the vocal folds, the model lacks a mechanism whereby the energy is lost in tissue. Including such effects will promote a balance between the losses that the tissue possesses and the input energy from the transglottal pressure gradient. In human phonation, the inclusion of material viscosity will have an effect on the vocal fold dynamics and indirectly an effect on the aerodynamics, due to the closely coupled nature of our model. The inclusion of viscosity is thus planned for the follow-up modeling using the PIB model.

The present work has used a very simple model of the vocal fold material consisting of a two-deep layer of particles with stretching, bending, and fixed point restorative forces between the particles, as well as a stretching force between the outer and inner layers of the material model. While the present work demonstrates the basic viability of such a simplified model of the material, it is clear that further work is needed to make a more realistic model of the vocal folds. Further, it is seen how even with this simplified model of the vocal folds, simulations are obtained which possess volume flows that exhibit many of the properties expected of a viable model of vocal fold dynamics and aerodynamics. Analysis of the 8 cm H_2O case and volume flow graphs for 2, 4, 8, and 12 cm H_2O cases have provided results suggesting a sensitivity of the model to the imposition of given external transglottal pressures.

The development is currently underway of a generalized PIB method which couples to an extended material rather than layers to the interface with the air. Additional models of the material will be supplied in subsequent studies. One advantage that the PIB method approach provides is the option to tune the material parameters to obtain qualitatively and quantitatively different dynamics. This aspect of the PIB model, wherein various "computational vocal fold materials" are designed and their properties explored is just beginning to be exploited. Here the stage has been set for the application of the PIB model to such questions.

ACKNOWLEDGMENTS

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**(6), 2003–3012.

Alipour, F., and Scherer, R. C. (**2000**). "Vocal fold bulging effects on phonation using a biophysical computer model," J. Voice **14**(4), 470–483.

Alipour, F., and Titze, I. (**1996**). "Combined simulation of two-dimensional airflow and vocal fold vibration," *Vocal Fold Physiology, Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher, Singular Publishing Group, Inc., San Diego, Chap. 2.

Alipour, F., and Scherer, R. C. (**2004**). "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am. **116**(3), 1710–1719.

Almgren, A. S., Bell, J. S., Colella, P., Howell, L. H., and Welcome, M. L. (**1998**). "A conservative adaptive projection method for the variable density incompressible Navier–Stokes equations," J. Comput. Phys. **142**, 1–46.

Cranen, B., and Boves, L. (**1985**). "Pressure measurement during speech production using semiconductor miniature pressure transducer: Impact on models for speech production," J. Acoust. Soc. Am. **77**, 1543–1551.

Kim, Y. (2003). "The penalty immersed boundary and its application to aerodynamics," Ph.D. thesis, New York University, Department of Mathematics (available at http://www.lib.umich.com/dissertation/individuals.html, order AAT 3105887).

McQueen, D. M., and Peskin, C. S. (**2001**). "Heart simulation by an immersed boundary method with formal second-order accuracy and reduced numerical viscosity," in Mechanics for a New Millennium, Proceedings of the International Conference on Theoretical and Applied Mechanics (ICTAM) 2000, edited by H. Aref and J. W. Phillips.

Peskin, C. S. (**1972**). "Flow patterns around heart valves: A digital computer method for solving the equations of motion," Ph.D. thesis, Albert Einstein College of Medicine, p. 211 (available at http://www.umi.com/hp/Products/DirExpress.html, order number 7230378).

Peskin, C. S., and McQueen, D. M. (**1989**). "A three-dimensional computational method for blood flow in the heart:(I) Immersed elastic fibers in a viscous incompressible fluid," J. Comput. Phys. **81**, 372–405.

Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**(2), 1249–1260.

Titze, I. R. (**1994**). *Principles of Voice Production*, Prentice Hall, Englewood Cliffs, NJ.

Titze, I. R., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.

Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese

Anders Löfqvist[a]

Haskins Laboratories, New Haven, Connecticut 06511 and Department of Logopedics,
Phoniatrics & Audiology Clinical Sciences, Lund, Lund University, Lund, Sweden

This paper examines the coordination of lip and tongue movements in sequences of vowel-bilabial consonant-vowel where the duration of the oral closure for the consonant is varied for linguistic purposes. Native speakers of Japanese served as subjects. The linguistic material consisted of Japanese word pairs that only differed in the duration of the labial consonant, which was either long or short. Recordings were made of lip and tongue movements using a magnetometer system. Results show a robust difference in closure duration between the long and short consonants. The tongue movement from the first to the second vowel had a longer duration in the long than in the short consonants, and its average speed was slower in the long consonant. The size of the tongue movement path between the vowels did not consistently differ between the long and short consonants. The tongue movement almost always started before the oral closure for the consonant, while the onset of the lip movement towards oral closure mostly started before that of the tongue movement. The offset of the tongue movement occurred after the release of the closure, but there was no clear pattern for the long and short consonants. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2345832]

Pages: 2872–2883

I. INTRODUCTION

This paper examines the coordination of lip and tongue movements in sequences of vowel-bilabial consonant-vowel, where the duration of the oral closure for the consonant is varied for linguistic purposes, using speakers of Japanese. An earlier study of American English (Löfqvist and Gracco, 1999) showed that the onset of the tongue movement from the first to the second vowel in such a VCV sequence almost always started before the oral closure for the consonant. In addition, more than 50% of the tongue movement trajectory between the vowels occurred during the oral closure. There was also a weak positive correlation between the magnitude of the tongue movement for the vowels and the interval between tongue movement onset and the onset of the lip closing movement. That is, the tongue started to move earlier than the lips as the tongue trajectory increased.

In Japanese, the ratio of closure duration for long and short consonants is about 2:1 (Beckman, 1982; Han, 1994; Hirata and Whiton, 2005). There is an extensive body of acoustic studies of the Japanese sound system with particular emphasis on the role of the mora for speech timing (e.g., Warner and Arai, 2001). The long consonants in Japanese are sometimes referred to as "geminates," and such a consonant contributes one mora; it is also traditionally assumed that a mora boundary occurs in the long consonant. This paper does not address the general issue of mora timing, however. Its primary focus is on speech motor control, capitalizing on the length distinction in Japanese sound structure to study coordination of lip and tongue movements under two conditions of consonant length. It is thus beyond the scope of this

study to review the many different views of the mora in Japanese. The reader is referred to Vance (1987) and Warner and Arai (2001).

If the duration of the oral closure for the consonant is increased, a speaker can in principle use two strategies for the coordination of the tongue and lip movements, as illustrated in the bottom panel of Fig. 1; the top panel in Fig. 1 shows tongue, jaw, and lip movements between the two vowels in the word /kami/ with a short labial consonant. One strategy would make the same tongue movement trajectory for a long and a short consonant. Such a strategy could have one or two versions, or a combination of them. The onset of the tongue movement might be shifted later relative to the oral closure, so that the tongue movement reached its position for the second vowel at the same point in time for the long and short consonant, pattern 1a in the lower panel of Fig. 1. Alternatively, the onset of the tongue movement could have the same temporal relationship to the consonant closure, pattern 1b in the lower panel of Fig. 1. In this second case, the interval between the offset of the tongue movement and the release of the oral closure for the labial consonant would thus increase for the long consonant. Pattern 1c in the lower panel of Fig. 1 shows a tongue movement trajectory that is shifted both relative to the onset and offset of the labial consonant. A second strategy would be to modify the tongue movement during the long consonant, so that the relative timing between the tongue movement for the vowels and the lip movements for the oral closure of the consonant would be more or less the same for the long and short consonants, pattern 2 in the lower panel of Fig. 1. This would imply making a slower tongue movement trajectory for the long than for the short consonant. The timing between the tongue and lip movements would thus be similar in the two

---

[a]Electronic mail: lofqvist@haskins.yale.edu
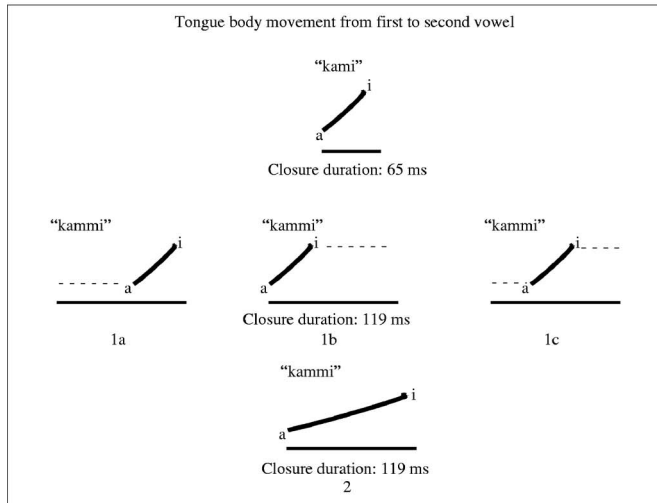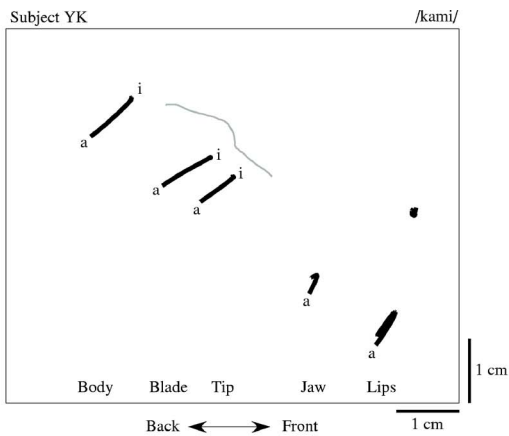
FIG. 1. The top panel shows tongue, jaw, and lip movements between the two vowels in the word /kami/ with a short labial consonant; the gray line is an outline of the hard palate. The bottom panel shows the actual tongue body movement between the two vowels in the word with a short labial consonant and potential tongue movement trajectories in a word with a long consonant. See the text for further explanation.

conditions of consonant length. Experimental evidence for the second strategy has been provided by (Smith, 1992, 1995). Using the x-ray microbeam system (Nadler *et al.*, 1987), she showed that Japanese speakers tended to increase articulatory intervals from the medial consonant to the second vowel in long compared to short consonants.

The present study extends the findings of Smith, using the measurement procedures developed by Löfqvist and Gracco (1999). For example, to analyze temporal patterns of tongue movements, Smith made separate measurements in the vertical and horizontal tongue movements, and also used "target intervals" based on a criterion of minimum movement. In this study, a single measure of tongue movement speed, taking into account both the vertical and horizontal movement components, is used for determining onset and offset of tongue movements. In addition, measurements are made of the duration and magnitude of the tongue movement trajectory, the average speed of the tongue movement between the two vowels, and the coordination of lip and tongue movements. The question addressed is whether these measures differ between words with long and short consonants. The hypothesis, based on the work of Smith, is that the

tongue movement in the words with a long consonant will be slower and of longer duration than in a word with a short consonant, implying that speakers maintain a similar overall phasing between lip and tongue movements in long and short consonants.

II. METHOD

A. Subjects

Five native speakers of Japanese, three male and two female, served as subjects. They reported no speech, language, or hearing problems. They were naive as to the purpose of the study. Before participating in the recording, they read and signed a consent form. (The experimental protocol was approved by the IRB at the Yale University School of Medicine.)

B. Linguistic material

The linguistic material consisted of Japanese words with a sequence of vowel-labial nasal-vowel. These words formed minimal pairs, where the only difference between the pairs was the duration of the labial consonant. The words were designed to require a substantial amount of tongue movement from the first to the second vowel. The following words were used: /kami, kammi/, /kamee, kammee/, /kema, kemma/, /semu, semmu/. The productions of the last word pair turned out to have a very small amount of tongue movement, so they were not analyzed. In addition, to examine tongue movements in symmetrical vowel contexts, the following words were also recorded for three of the subjects: /kama, kamma/, /ʃimi, ʃimmi/. The linguistic material was organized into randomized lists and presented to the subjects in Japanese writing, with the words occurring in a short frame sentence. Fifty repetitions of each word were recorded.

C. Movement recording

The movements of the lips, the tongue, and jaw were recorded using a three-transmitter magnetometer system (Perkell *et al.*, 1992); when proper care is taken during the calibration, the spatial resolution of the system is on the order of 0.5 mm. Receivers were placed on the vermilion border of the upper and lower lip, on three positions of the tongue, referred to as tip, blade, and body, and on the lower incisors at the gum line. Two additional receivers placed on the nose and the upper incisors were used for the correction of head movements. The lip and jaw receivers were attached using Isodent, a dental adhesive, while the tongue receivers were attached using Ketac-Bond, another dental adhesive. Care was taken during each receiver placement to ensure that it was positioned at the midline, with its long axis perpendicular to the sagittal plane. Two receivers attached to a plate were used to record the occlusal plane by having the subject bite on the plate during the recording. All data were subsequently corrected for head movements and rotated to bring the occlusal plane into coincidence with the *x* axis. This rotation was performed to obtain a uniform coordinate system for all subjects (cf. Westbury, 1994).
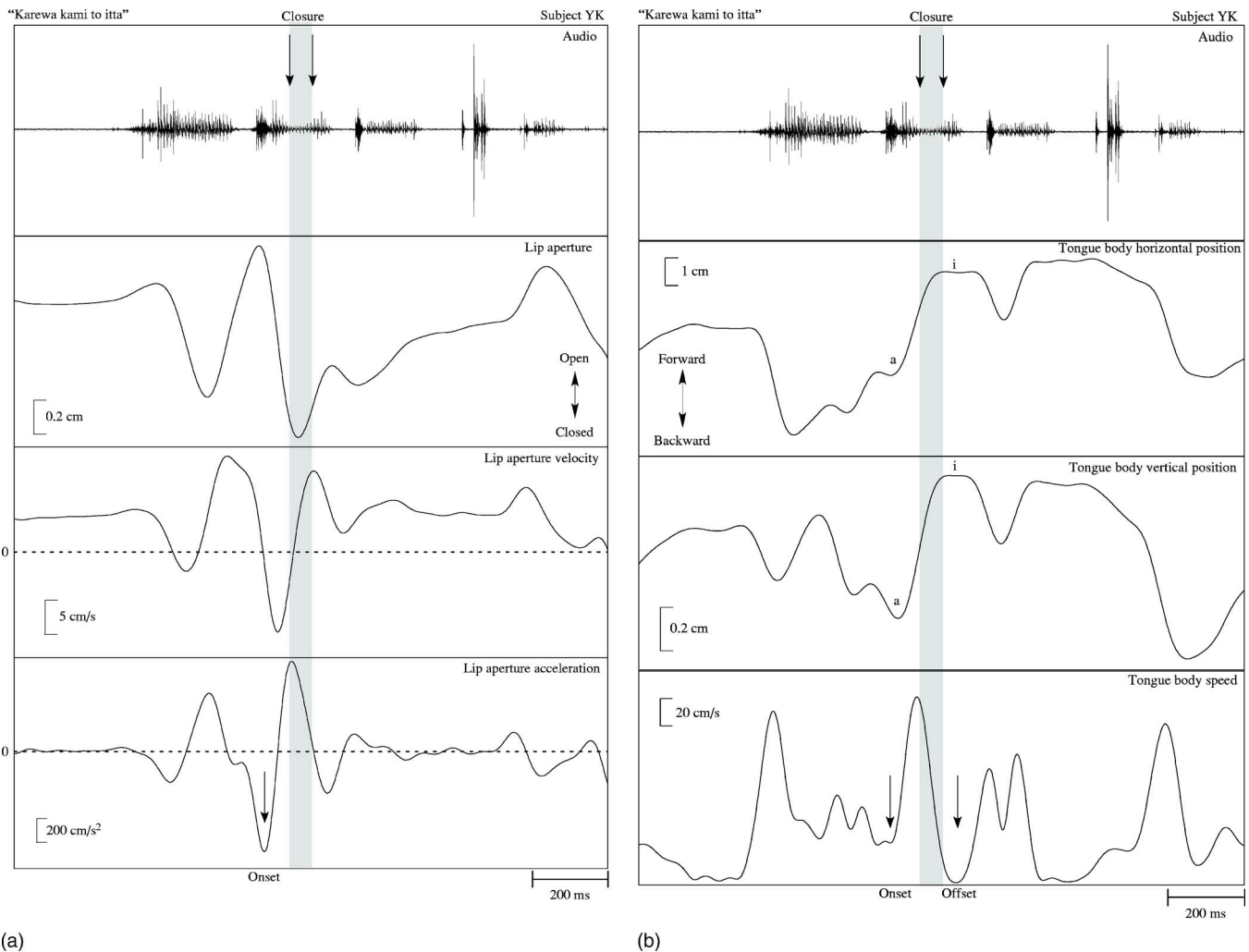
FIG. 2. (a) Audio and lip aperture signals for the word /kami/, showing the point in the acceleration signal used for defining the onset of the lip closing movement. (b) Audio and tongue body signals for the word /kami/, showing the points used in the speed signal for defining the onset and offset of the tongue movement between the two vowels. The baseline in the bottom panel with the speed signal represents zero speed.

The articulatory movement signals (induced voltages from the receiver coils) were sampled at 500 Hz after low-pass filtering at 200 Hz. The resolution for all signals was 12 bits. After voltage-to-distance conversion, the movement signals were low-pass filtered using a 25-point triangular window with a 3-dB cutoff at 14 Hz; this was done forwards and backwards to maintain phase. A measure of lip aperture was obtained by calculating the vertical difference between the upper and lower lip receivers. To obtain instantaneous velocity of the lip aperture and tongue receivers, the first derivative of the position signals was calculated using a 3-point central difference algorithm. For the lip aperture signal, its acceleration was also calculated from its velocity. For each tongue receiver, its speed $v = \sqrt{(\dot{x}2 + \dot{y}2)}$ was also calculated. The velocity signals were smoothed using the same triangular window. All the signal processing was done using the Haskins Analysis Display and Experiment System (HADES) (Rubin and Löfqvist, 1996). The acoustic signal was pre-emphasized, low-pass filtered at 4.5 kHz, and sampled at 10 kHz.

The onset and release of the oral closure for the nasal consonant were identified in waveform and spectrogram displays of the acoustic signal. They were both identified by a change in the amplitude and the spectral properties.

To define the onset of the closing movement of the lips for the nasal consonant, an algorithmically obtained minimum in the lip aperture acceleration signal just before oral closure was used, following Löfqvist and Gracco (1999); see Fig. 2(a). In the words with asymmetrical vowel contexts, tongue movement onsets and offsets were defined algorithmically in the tongue body speed signal as minima during, or close to, the first and second vowels; see Fig. 2(b). We should note that, at these points in time, the horizontal and vertical velocity of the tongue is usually not zero. The duration of the tongue body movement was measured between its onset and offset. The magnitude of the tongue movement trajectory from the first to the second vowel was obtained by summing the Euclidean distances between successive samples of the tongue body vertical and horizontal receiver positions from movement onset to movement offset. The average speed of the tongue body was obtained by adding the speed of all the individual samples between movement onset and offset and then dividing by the number of samples in the interval. The vertical and horizontal positions of the tongue

body receiver at movement onset and offset were measured to examine if any observed changes in movement magnitude were due to changes in onset position, offset position, or both. Three temporal measurements were made. The first one consisted of the interval between tongue movement onset and oral closure; it provides information about the tongue movement relative to consonant closure. The second one was the interval between tongue movement onset and lip closing movement onset; it examines the coordination between tongue and lip movements. The third one was the interval between tongue movement offset and oral release; it is useful for examining the tongue movement relative to consonant release. The kinematic signals are expressed in a maxilla-based coordinate system. Thus, the tongue body movement includes the contribution of the jaw, which is appropriate when we are interested in the tongue as the end effector.

In the words with symmetrical vowel contexts, no temporal landmarks could be defined in the tongue movement signals. For these words, the movement path of the tongue, and the jaw were measured during the acoustically defined oral closure for the nasal consonant, and also the tongue body and jaw positions at consonant onset and offset. The jaw was included, since it might contribute to the labial closure, and hence to tongue movements, in particular in the words with the open vowel /a/.

To assess differences between the long and short consonants for each subject, T-tests were used. Given the large number of comparisons, a conservative α-level of 0.001 was adopted. Since the variances usually differed between the long and short consonants, as shown by Levene's test, the statistical tests assumed unequal variances, so the degrees of freedom were adjusted (e.g., Winer *et al.*, 1991, p. 67).

III. RESULTS

A. Consonant duration

The duration of the oral closure for the long and short nasal consonants is summarized in Fig. 3. For all speakers, there is a clear and robust difference between the long and short consonants, with the long ones having about twice the duration of the short ones. The statistical analysis showed all the differences to be significant (for /kami, kammi/ $t=$ -20.28, -24.39, -24.83, -18.28, and -17.86, for speakers YK, YM, KN, SO, and TT, respectively; for /kamee, kammee/ $t=-19.88$, -29.19, -40.79, -38.19, and -23.01, and for /kema, kemma/ $t=-19.72$, -28.78, -38.73, -26.32, and -15.67, with $p<0.001$ in all cases).

B. Tongue movements

The first analysis focused on the kinematic properties of the tongue movement from the first to the second vowel.

1. Tongue body movement duration

Figure 4 shows the duration of the tongue body movement from the first to the second vowel. For most subjects and words, the duration is longer in the long than in the short consonant. The two exceptions were for the productions of
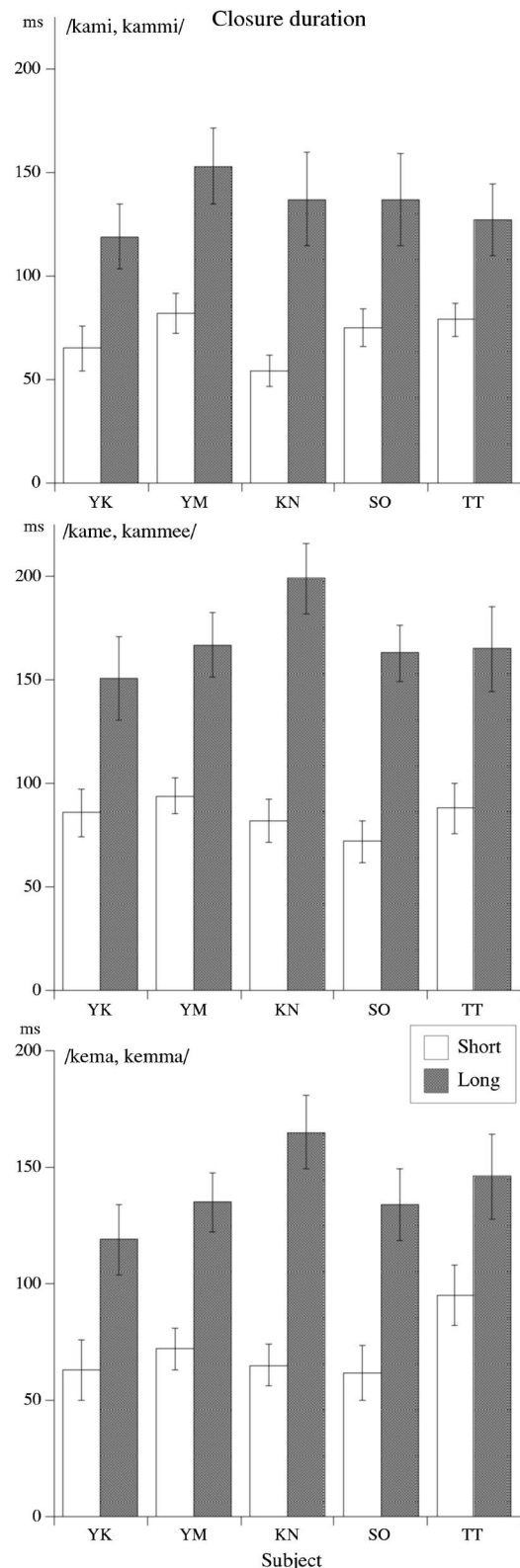
FIG. 3. Closure duration of the bilabial nasal consonant (mean and standard deviation).

/kema, kemma/ by speakers SO and TT, where the duration of the movement trajectory was longer for the short consonant, but the difference was not significant in either case (t $=1.27$, and 3.23 for SO and TT, ns). For all the other cases, the difference was significant (for /kami, kammi/ $t=-11.2$, -8.19, -4.2, -12.17, -13.63; for speakers YK, YM, KN, SO,

Duration of tongue body movement from the first to the second vowel

FIG. 4. Duration of the tongue body movement from the first to the second vowel (mean and standard deviation).

Average speed of the tongue body movement from the first to the second vowel

FIG. 5. Average speed of the tongue body movement from the first to the second vowel (mean and standard deviation).

2. Average speed of tongue body

Figure 5 shows the average speed of the tongue body movement from the first to the second vowel. Overall, the average speed is slower for the long than the short consonant. With the exception of /kami, kammi/ of speaker YM ($t=3.2$, ns) the difference was statistically significant (for /kami, kammi/ $t=6.39$ 14.91. 7.0, 8.64, for speakers YK, KN, SO, and TT, respectively; for /kamee, kammee/ $t=9.55$, 6.63, 22.12, 10.61, 13.42; and for /kema, kemma/ $t=-6.73$,

and TT, respectively; for /kamee, kammee/ $t=-13.48$, -7.55, -10.34, -17.58, -14.55; and for /kema, kemma/ $t=-9.97$, -4.64, -4.97, for speakers YK, YM, and KN, with $p < 0.001$ in all cases).

Anders Lofqvist: Interarticulator programming

Length of the path of the tongue body movement from the first to the second vowel
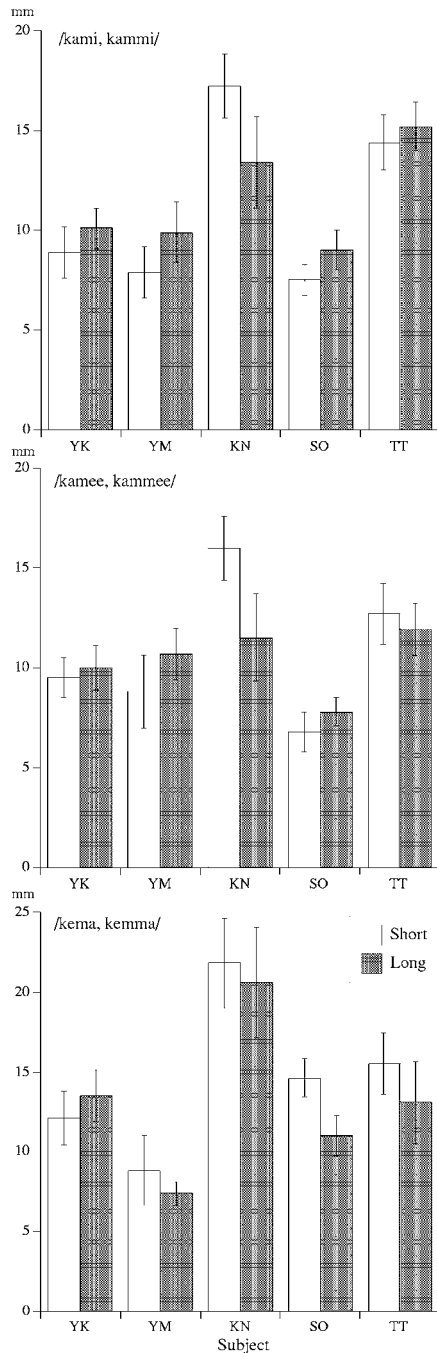
FIG. 6. Path of the tongue body movement from the first to the second vowel (mean and standard deviation).

17.56, 12.52, 11.91, 3.84, for speakers YK, YM, KN, SO, and TT, respectively, with $p < 0.001$ in all cases).

3. Magnitude of tongue movement path

Figure 6 shows the path of the tongue body movement from the first to the second vowel. Here, the results differ both within and across speakers. Speaker YK always produced the long consonant with longer movement path, but the difference was only statistically significant for /kami, kammi/ ($t=-5.17$, $p < 0.001$) and /kema, kemma/ ($t=-4.3$, $p < 0.001$) but not for /kamee, kammee/ ($t=-1.98$, ns). Speaker YM made significant differences between all the

word pairs, but in /kami, kammi/ ($t=-7.26$, $p < 0.001$) and /kamee, kammee/ ($t=-5.86$, $p < 0.001$) the path was longer for the long consonant, whereas the opposite pattern was found for /kema, kemma/ ($t=4.49$, $p < 0.001$). Speaker KN consistently produced the short consonant with a longer tongue movement path, but the difference was only significant in two cases ($t=9.38$ and 11.6 for /kami, kammi/ and /kamee, kammee/, respectively, with $p < 0.001$ in both cases, but not for /kema, kemma/ $t=1.97$, ns). Also for speaker SO, the results varied between words. For /kami, kammi/ and /kamee, kammee/ the path was longer for the long than for the short consonant ($t=-8.53$, and -5.29, with $p < 0.001$ in both cases). However, the opposite was found for /kema, kemma/ ($t=14.03$, $p < 0.001$). Finally, speaker TT only showed a significantly longer path for the short consonant in /kema, kemma/ ($t=5.08$, $p < 0.001$), while no difference was found for the other two words ($t=-3.34$, and 2.83, ns, for /kami, kammi/ and /kamee, kammee/, respectively).

The path of the tongue body movement was not influenced by consonant length in a systematic manner across speakers. Whenever an influence was found, it could be due to a difference in onset position, offset position, or both, but there was no consistent pattern across subjects.

In summary, the results for the tongue movement from the first to the second vowel show that it generally had a longer duration for the long than the short consonant. Its average speed was always lower for the long than for the short consonant. The path of the tongue movement showed no consistent influence of consonant length.

C. Interarticulator programming

The next analysis examined the timing between the lip and tongue movements for the labial consonant and the two vowels. Three different measurements were made.

1. Onset of tongue body movement relative to onset of oral closure

Figure 7 summarizes the duration of the interval between tongue movement onset and oral closure for the consonant. Of note here is that, with the exception of only one case, /kammee/ for subject KN, the tongue movement onset always occurs before the closure. Overall there is no clear pattern for the long and short consonants. For /kami, kammi/, only subjects KN and YM showed a reliable difference (KN, $t=11.1$, $p < 0.001$, and YM, $t=12.9$, $p < 0.001$), but they were not in the same direction. The corresponding t values for the other subjects were 1.45, 0.36, and 0.83 for YK, SO, and TT, respectively, all ns. Also for the pair /kamee, kammee/, the results were inconsistent. Subjects YM, KN, and SO had significant differences ($t=11.30$, -10.41, and 3.39, with $p < 0.001$ in all cases), but only speakers YM and SO showed this interval to be longer for the long than for the short interval. Speaker KN produced the long consonant with a tongue movement that started after the closure. The other two speakers, YK and TT, showed no difference ($t=2.75$, and -2.37, ns). In contrast, four of the speakers produced the long consonant in /kema, kemma/ with an earlier start of the tongue movement for the short consonant ($t=-4.45$, -9.74,

Interval between the onset of the tongue body movement
from the first to the second vowel and the oral closure

FIG. 7. Interval between the onset of the tongue body movement from the
first to the second vowel and the oral closure (mean and standard deviation).

Interval between the onset of the tongue body movement from the first
to the second vowel and the onset of lip movement for closure
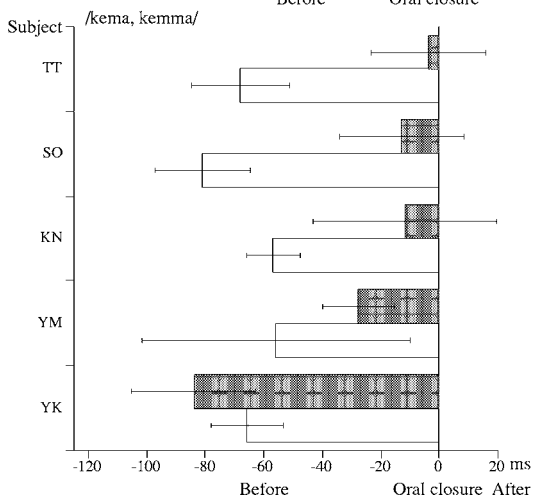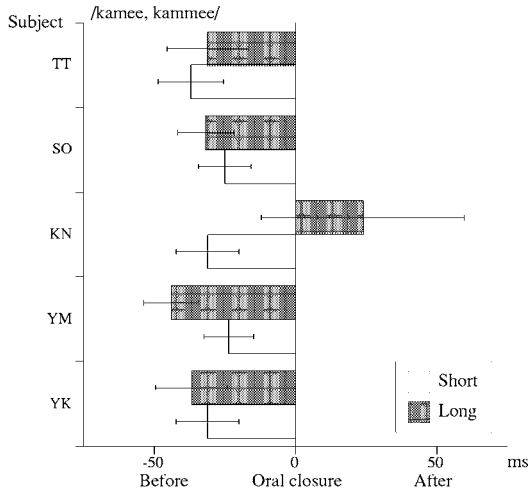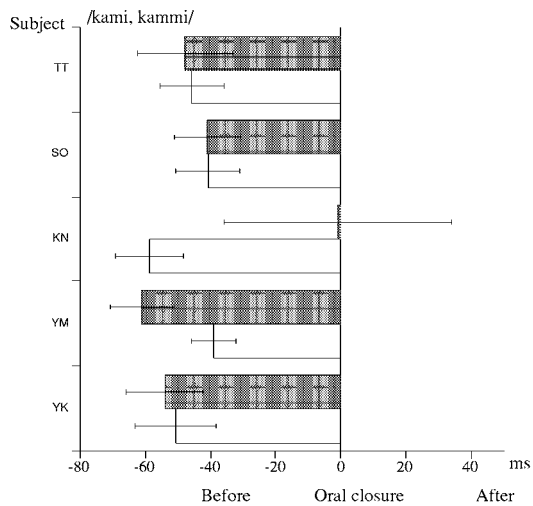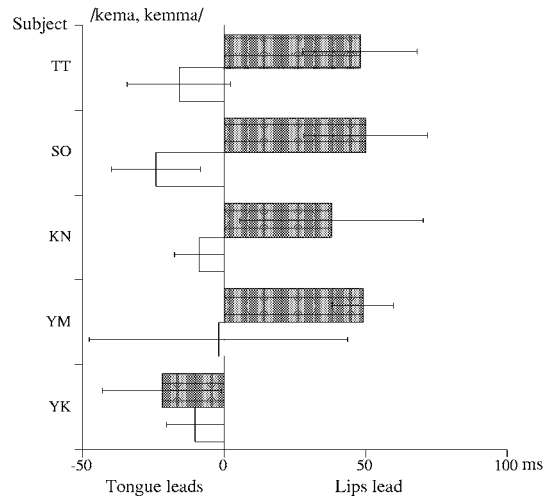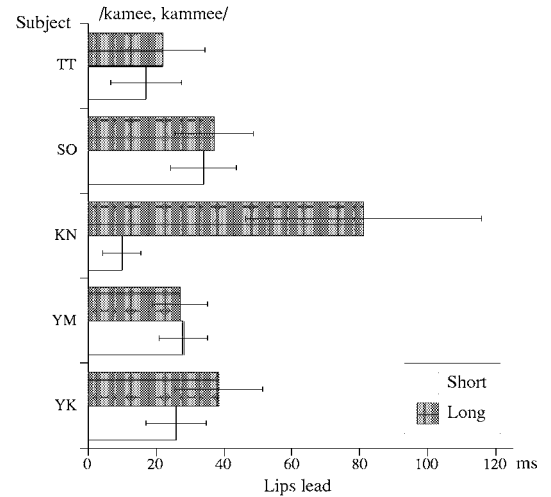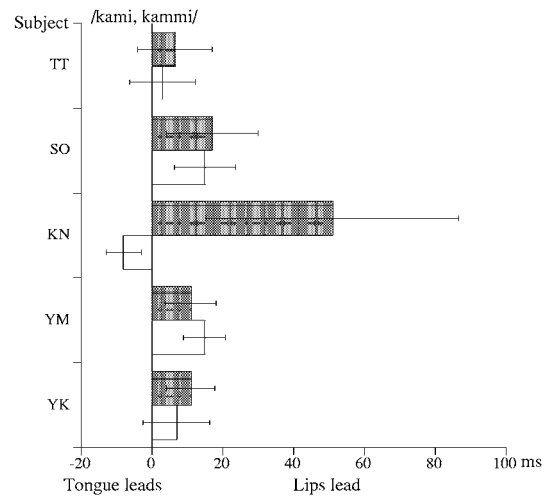
FIG. 8. Interval between the onset of the tongue body movement from the
first to the second vowel and the onset of the lip movement towards closure
(mean and standard deviation).

−17.87, and −17.59 for YM, KN, SO, and TT, respectively,
with $p < 0.001$ in all cases). Speaker YK showed the oppo-
site pattern($t = 5.26$, $p < 0.001$).

2. Onset of tongue body movement relative to onset of lip movement for closure

The results for this interval are shown in Fig. 8. The
overall results suggest that the lips lead, but there is no clear

difference between the long and short consonants. For /kami,
kammi/, speakers YK, YM, and KN showed a significant
difference between the long and short consonants ($t = -2.61$,
3.57, and −11.51, for YK, YM, and KN, respectively, with
$p < 0.001$). However, the difference went in opposite direc-
tions for YK and KN, compared to YM. For the remaining

two speakers, SO and TT, there was no difference ($t=0.92$ and -1.93, ns). For /kame, kammee/, only speakers YK and KN had a reliable difference in this interval between the long and short consonants ($t=-5.59$, and 14.36, $p<0.001$) and here the lips led with a longer time for the long than for the short consonants. For the other speakers, YM, SO, and TT, no difference was found ($t=-0.67$, -1.54, and -2.04, ns). Four of the subjects, YM, KN, SO, and TT showed a longer lead of the lip movement in /kemma/ than in /kema/ ($t=-7.65$, 9.78, -19.6, and -16.64, for YM, KN, SO, and TT, respectively, with $p<0.001$ in all cases). Note that, for the short consonant /kema/, these four subjects started the tongue movement before the lip movement. Subject YK showed a different pattern with the tongue movement leading the lip movement, with a significant longer tongue movement lead for the long consonant ($t=3.49$, $p<0.001$).

3. Offset of tongue body movement relative to oral release

As shown in Fig. 7, the onset of the tongue movement from the first to the second vowel almost always started before the oral closure for the consonants. Its offset tended to occur after the oral release; see Fig. 9. The pattern was, however, quite variable as shown by the standard deviations. Only speakers YK and YM showed a significant difference between the long and short consonants in /kami, kammi/, with the tongue movement offset occurring later for the short consonant ($t=3.19$ and 4.06, $p<0.001$). The remaining three subjects had no difference ($t=0.06$, 2.39, and 1.13, for speakers KN, SO, and TT, respectively). A similar pattern with a later offset for the short than for the long consonants in /kame, kammee/ was found for speakers YK, SO, and TT, but the difference was only significant for SO and TT ($t=5.03$, and 4.35, $p<0.001$), and not for YK ($t=2.54$, ns). The remaining two speakers, YM and KN, showed no difference ($t=0.09$, and -1.28, ns). Finally, for the pair /kema, kemma/, only subjects YK and SO showed a reliably later offset in the short than in the long consonant ($t=5.3$, and 3.59, $p<0.001$). The other subjects showed no difference ($t=0.88$, 1.89, and -0.6, ns, for YM, KN, and TT, respectively).

4. Symmetrical vowel contexts

Two word pairs with symmetrical vowel contexts, /kama, kamma/, /ʃimi/, /ʃimmi/, were examined in three of the subjects. The results for tongue body and jaw path lengths during the acoustically defined oral closure are shown in Fig. 10. For the open vowels in /kama, kamma/, the movement path during the oral closure is longer for the long than for the short consonant for both the tongue body and the jaw (tongue body: $t=-12.88$, -5.94, and -11.73, for subjects YK, SO, and TT, with $p<0.001$; jaw $t=-16.8$, -22.78, and -11.58, $p<0.001$). For the high vowels in /ʃimi, ʃimmi/, the path of the tongue body movement is longer during the short than during the long consonant ($t=6.51$, 6.51, and 8.17,
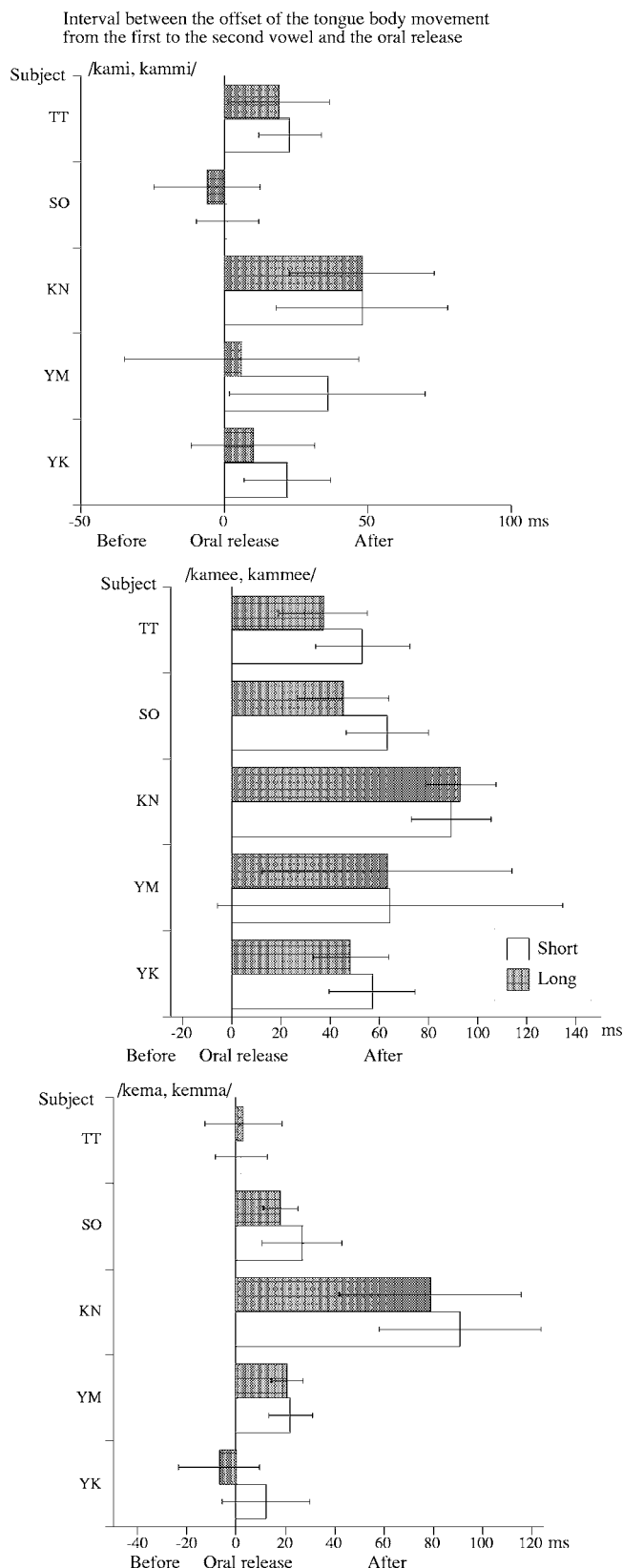
FIG. 9. Interval between the offset of the tongue body movement from the first to the second vowel and the oral release (mean and standard deviation).

$p<0.001$). In contrast, the path of the jaw during the oral closure is longer for the long than for the short consonant, but the difference was only statistically significant for subjects YK and TT ($t=-4.37$, and -6.83, $p<0.001$), but not for

Length of the path of the tongue body movement during oral closure

Length of the path of the jaw movement during oral closure
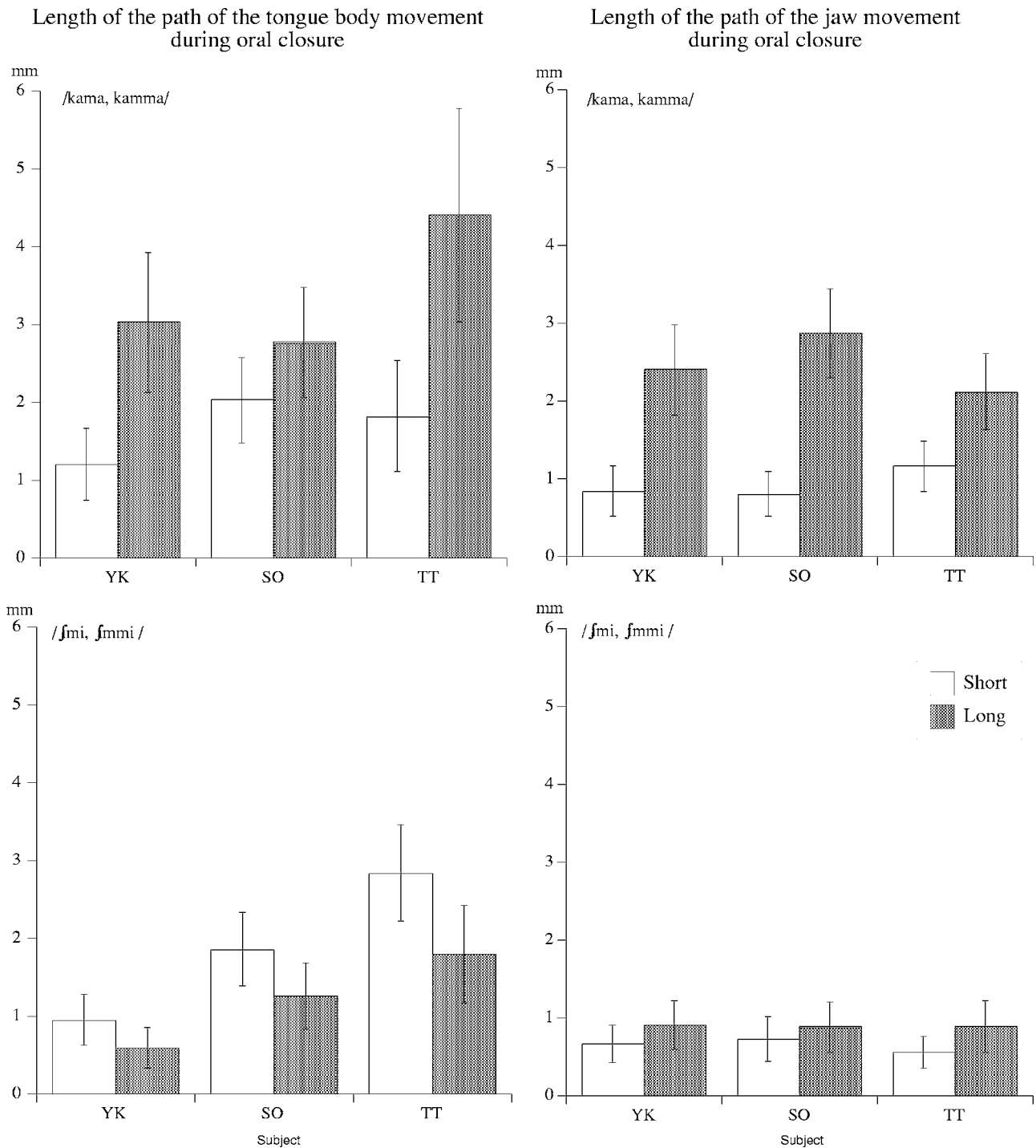
FIG. 10. Paths of the tongue body and jaw movements during the oral closure (mean and standard deviation).

subject SO ($t=-2.55$, ns). The movement of the jaw is made to assist in making the labial closure, which is reflected in the longer movement path of the jaw for the open vowel context /a/ than for the high vowel context /i/. As shown in the lower panels of Fig. 11, the jaw positions at the onset and offset of the nasal consonant do not differ very much in either vowel context. Nor is there any clear pattern within and across subjects for the long and short consonants.

Although the tongue is not rigidly coupled to the jaw, some of its movement is related to that of the jaw. This was particularly the case during the oral closure in the open

vowel context, where the tongue body positions at closure and release were almost identical for subject YK, while the offset position tended to be lower for subjects SO and TT (Fig. 11, upper left panel). In the high vowel context, however, the tongue movement appeared to be more related to coarticulatory influences, since the tongue body had moved forward and upward during the oral closure; the movement magnitude ranged from 0.5 to 1.0 cm in both the horizontal and vertical dimensions (Fig. 11, upper right panel). This movement was most likely made to position to the tongue for the upcoming /t/ in the carrier phrase.

Tongue body and jaw position at consonant onset and offset in symmetrical vowel contexts
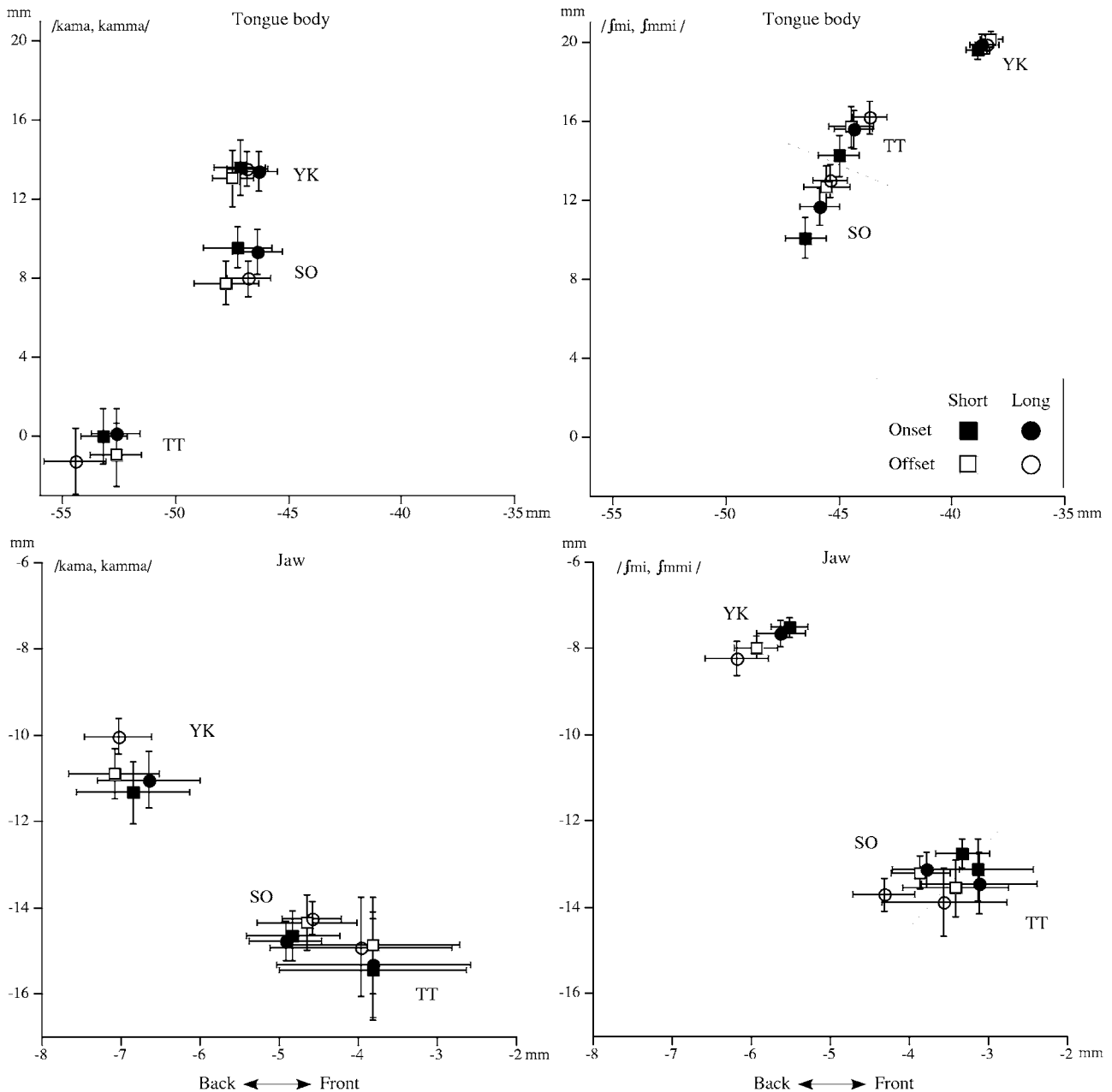
FIG. 11. Tongue body and jaw positions at onset and offset of the nasal consonant in symmetrical vowel contexts (mean and standard deviation). For clarity of presentation, the horizontal and vertical scales are different. The dashed lines in the two right panels separate the data points for subjects SO and TT.

IV. DISCUSSION

This study examined the influence of consonant duration on the coordination of lip and tongue movements in VCV sequences with a labial nasal consonant. It was hypothesized that a speaker could use one of two possible strategies of interarticulator programming if the duration of the consonant was increased. In one of them, the tongue movement trajectory would be similar for the long and short consonants; thus, the timing between the lip and tongue movements would change. In the second one, the tongue movement would be modified for the long consonant, resulting in a similar coordination of lip and tongue movements for both the long and short consonants. The present results clearly support the second strategy.

They thus show that speakers of Japanese modify their tongue movements when they produce a sequence of a vowel-bilabial consonant-vowel with a long and a short consonant. In particular, the duration of the tongue movement is increased, while its average speed is reduced in the long consonant. However, the path of the tongue movement did not vary consistently with vowel duration. The onset of the tongue movement started before the oral closure for both the long and short consonants. Similarly, the offset of the tongue movement occurred after the release of the consonant. Thus, the coordination of the tongue movement relative to the onset and release of the oral closure was similar for the long and short consonants. Importantly, all five speakers showed the same modification of tongue movement speed and dura-

tion for the long consonant. These findings are similar to the ones reported by (Smith, 1992, 1995). They thus suggest that the Japanese speakers tend to maintain a similar, but not identical, coordination between the tongue and lip movements across consonant length.

Although all subjects changed the tongue movement between the two vowels, they did not change it in such a way that its duration in the long consonant was about twice the duration of the movement for the short consonant; see Fig. 4. Thus, the duration of the tongue movement did not correlate closely with the duration of the oral closure. The onset and offset of the tongue movement were defined as minima in the speed signal, but as suggested by the large variability in some of the temporal intervals in Figs. 7–9, these points need not be under tight temporal or spatial control. Tongue positions for vowels do not have point targets, but rather regions, that are influenced by phonetic context and speaking rate (e.g., Gay, 1974; Guenther, 1995).

As shown in Fig. 5, all subjects reduced the speed of the tongue body movement between the two vowels in the long consonant, thus maintaining a similar interarticulator timing for the two consonant durations. However, no subject completely stopped the movement of the tongue. This finding could follow from a general principle of motor control, based on a cost minimization, often expressed as a minimum jerk criterion (Flash and Hogan, 1985; Flash, 1987) or as a smoothness constraint (Uno *et al.*, 1989). Thus, speakers avoid excessive accelerations and decelerations of the tongue by keeping it moving.

The present results are generally in agreement with those of Löfqvist and Gracco (1999). As in their study, the onset of the tongue movement very rarely starts after the oral closure. They also found that the tongue started to move earlier than the lips as the tongue trajectory increased (their Fig. 7). A similar analysis of the coordination of tongue and lip movement onsets and tongue movement magnitude in the present results showed the same overall trend for all the speakers, although the correlations were generally less than 0.5. Interestingly, an inspection of Figs. 6 and 8 shows that there is sometimes a relationship between the path of the tongue movement and the interval between tongue movement onset and lip movement onset for the individual speakers. For example, for /kema, kemma/, speakers YM, KN, SO, and TT all had a longer path for the short than for the long consonant. These subjects also had the tongue movement leading the lip movement in producing the short consonant. In addition, speaker YK had a longer movement path for the long consonant, and also an earlier tongue movement onset for the long consonant. Similarly, speaker KN consistently produced a longer path in for the short than for the long consonant (Fig. 6) and also had a shorter interval between tongue and lip movement onsets for the short consonant.

The present results thus show that speakers of American English and Japanese have similar coordination between lip and tongue movements in VCV sequences, and that this is true for both the long and short consonants in Japanese. At the same time, American English and Japanese are described to have stress- and mora timing, respectively, but the inter-

articulator programming of lip and tongue movements appears to be very similar for the two languages.

This raises the issue of whether the prosodic structure of a language will influence patterns of interarticulator timing. The results presented by Smith (1992, 1995) suggest that this may he the case, since speakers of Japanese and Italian, which also has a length contrast, may have different patterns of coordination in VCV sequences when the consonant length differs. Japanese is described as having mora timing, Italian is said to have syllable timing, but how, and if, such a difference will influence speech timing is debatable (e.g., Dauer, 1983; Bertran, 1999). Vatikiotis-Bateson and Kelso (1993) examined lower lip movements in English, French, and Japanese, using reiterant speech. Analyzing movement duration, amplitude, and peak velocity, they found that movement dynamics were qualitatively similar in the three languages but also that there were quantitative differences. Among these were speaking rate, the quantitative relationship between movement amplitude and peak velocity, and movement variability. Since the results presented by Smith for the Italian speakers were not as robust as for the Japanese speakers, it is of interest to further pursue this issue by applying the methodology used in the present study to languages such as Italian and Finnish.

Beckman, M. (**1982**). "Segment duration and the mora in Japanese," Phonetica **39**, 113–135.

Bertran, A. (**1999**). "Prosodic Typology: On the dichotomy between stress-timed and syllable-timed languages," Language Design **2**, 103–130.

Dauer, R. (**1983**). "Stress timing and syllable timing reanalyzed," J. Phonetics **11**, 51–62.

Flash, T. (**1987**). "The control of hand equilibrium trajectories in multi-joint arm movements," Biol. Cybern. **57**, 257–274.

Flash, T., and Hogan, N. (**1985**). "The coordination of arm movements: An experimentally confirmed mathematical model," J. Neurosci. **5**, 1688–1703.

Gay, T. J. (**1974**). "A cinefluorographic study of vowel production," J. Phonetics **2**, 255–266.

Guenther, F. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural betwork model of speech production," Psychol. Rev. **102**, 594–621.

Han, M. (**1994**). "Acoustic manifestations of mora timing in Japanese," J. Acoust. Soc. Am. **96**, 73–82.

Hirata, Y., and Whiton, J. (**2005**). "Effects of speaking rate on the single/geminate stop distinction in Japanese," J. Acoust. Soc. Am. **118**, 1647–1660.

Löfqvist, A., and Gracco, V. (**1999**). "Interarticulator programming in VCV sequences: Lip and tongue movements," J. Acoust. Soc. Am. **105**, 1864–1876.

Nadler, R., Abbs, J., and Fujimura, O. (**1987**). "Speech movement research using the new x-ray microbeam system," in Proceedings of the 11th Int. Congr. Phonetic Sci. (Academy of Sciences of the Estonian SSR, Tallinn), Vol. **1**, pp. 221–224.

Perkell, J. M., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.

Rubin, P. E. R., and Löfqvist, A. (1996). "HADES (Haskins Analysis Dis-

play and Experiment System)," Haskins Labs. Status Rep. Speech Res. (available at: http://www.haskins.yale.edu/MISC/DOCS/HADESRLn2.pdf)

Smith, C. (1992). "The Timing of Vowel and Consonant Gestures," Unpublished doctoral dissertation, Yale University.

Smith, C. (**1995**). "Prosodic patterns in the coordination of vowel and consonant gestures," *in Laboratory Phonology IV: Phonology and Phonetic Evidence*, edited by B. Connell and C. Arvaniti, (Cambridge University Press, Cambridge), pp. 205–222.

Uno, Y., Kawato, M., and Suzuki, R. (**1989**). "Formation and control of optimal trajectory in human multi-joint arm movement—Minimum torque change model," Biol. Cybern. **61**, 89–101.

Vance, T. (**1987**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY).

Vatikiotis-Bateson, E., and Kelso, J. A. S. (**1993**). "Rhythm type and articulatory dynamics in English, French and Japanese," J. Phonetics **21**, 231–265.

Warner, N., and Arai, T. (2001). "Japanese mora-timing: A review," **58**, 1–25.

Westbury, J. (**1994**). "On coordinate systems and the representation of articulatory movements," J. Acoust. Soc. Am. **95**, 2271–2273.

Winer, B., Brown, R., and Michels, K. (**1991**). *Statistical Principles in Experimental Design* (McGraw-Hill, Boston).

Is fundamental frequency a cue to aspiration in initial stops?

Alexander L. Francis,[a] Valter Ciocca, Virginia Ka Man Wong, and Jess Ka Lam Chan
Department of Speech and Hearing Sciences, University of Hong Kong, Hong Kong

One production and one perception experiment were conducted to investigate the interaction of consonant voicing and fundamental frequency at the onset of voicing (onset f_0) in Cantonese, a tonal language. Consonantal voicing in English can affect onset f_0 up to 100 ms after voicing onset, but existing research provides inconclusive information regarding the effects of voicing on f_0 in tonal languages where f_0 variability is constrained by the demands of the lexical tone system. Previous research on consonantal effects on onset f_0 provides two contrasting theories: These effects may be automatic, resulting from physiological constraints inherent to the speech production mechanism or they may be controlled, produced as part of a process of cue enhancement for the perception of laryngeal contrasts. Results of experiment 1 showed that consonant aspiration affects onset f_0 in Cantonese only within the first 10 ms following voicing onset, comparable to results for other tonal languages. Experiment 2 showed that Cantonese listeners can use differences in onset f_0 to cue perception of the voicing contrast, but the minimum extent of f_0 perturbation necessary for this is greater than is found in Cantonese production, and comparable to that observed in acoustic studies of nontonal languages. These results suggest that consonantal effects on onset f_0 are at least partially controlled by talkers, but that their role in the perception of voicing/aspiration may be a consequence of language independent properties of audition rather than listeners' experience with the phonological contrasts of a specific language.
© 2006 Acoustical Society of America. [DOI: 10.1121/1.2346131]

I. INTRODUCTION

In English, fundamental frequency at the vowel onset (onset f_0) is correlated with the phonological feature of voicing in initial stops, such that the onset f_0 following voiceless stops is higher than after voiced stops (Haggard, Ambler, and Callow, 1970; Hombert, 1978; House and Fairbanks, 1953; Lehiste and Peterson, 1961; Löfqvist, Baer, McGarr, and Seider Story, 1989; Ohde, 1984; Whalen, 1990). It has been argued that this pattern is *intrinsic* to voicing differences, meaning that the onset f_0 is determined by physiological and/or aerodynamic factors related to the relative timing of voicing and oral closure release in a manner that is not (entirely) under the control of the speaker (cf. Hombert, 1978; Ohala, 1978 pp. 25–29). However, the duration of consonant-related f_0 perturbations in English extends farther into the vowel than would be predicted by simple physiological or aerodynamic factors (Hombert, 1978). Because of its longer-than-expected duration, Kingston and Diehl (1994) argued that this effect is an intentional maneuver on the part of English speakers to help cue the perception of voiced consonants.

Further support for the hypothesis that English speakers may intentionally exaggerate consonant-related f_0 perturbations in order to enhance voicing perception is found in tonal languages: Gandour (1974) and Hombert (1977) both determined that voicing-related onset f_0 perturbations were shorter in Thai and Yoruba, respectively, than in nontonal languages, extending approximately 30–50 ms into the vowel. By contrast, onset f_0 perturbations are observed for 100 ms or more in nontonal languages such as English (House and Fairbanks, 1953; Lehiste and Peterson, 1961; Löfqvist, Baer, McGarr, and Seider Story, 1989; Whalen, 1990), French (Hombert, 1978) and Dutch (Löfqvist *et al.* 1989). Thus, it has been argued that tonal language speakers either actively seek to inhibit the extent of an intrinsic effect on f_0 or choose not to manipulate pitch as a voicing cue because they are already using pitch to cue lexical tone contrasts. In either case, the essential point is the same: Pitch is a strong cue for tone categorization but a relatively weak one for consonant identification (Abramson and Lisker, 1985; Haggard, Ambler, and Callow, 1970). Therefore, tone language speakers avoid allowing a consonant-associated f_0 event to affect pitch (and therefore tone) except perhaps very locally within the onset of a syllable (Hombert, 1978, p. 83). If tonal language speakers are able to control the effect of consonant voicing on onset f_0, then this effect must be at least partially controllable, and not purely intrinsic to the production of voicing.

However, one crucial causal link in this chain of reasoning is still missing: It is not clear whether tone language speakers are able to use f_0 as a cue to consonant voicing (or aspiration), let alone whether they actually do so in their native language. It has been shown that f_0 is not always a strong cue to consonant voicing even for speakers of languages that do show considerable correlation between onset f_0 and voice onset time (VOT) in production (Francis and

[a] Author to whom correspondence should be addressed. Department of Speech, Language and Hearing Sciences, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, Indiana 47907. Electronic mail: francisa@purdue.edu

Nusbaum, 2002; Haggard *et al.*, 1970; Whalen, Abramson, Lisker, and Mody, 1990). If f_0 is so critical to tone perception that speakers cannot afford to allow it to vary (much) according to voicing, then perhaps tone language listeners simply disregard this property of the signal when making consonant voicing decisions in the first place. In the present paper we revisit the interaction of consonantal perturbations of f_0 and tone and extend this research into the domain of perception. We ask whether native speakers of Cantonese, a tone language, produce consonants with limited voicing-associated f_0 perturbations like Thai and Yoruba speakers, and, if so, whether they are still able to use onset f_0 as a cue to consonant voicing.

A. Voicing and onset f_0

To some extent, the effects of English stop consonant voicing on onset f_0 are unexpected. Although English stops are often considered to contrast in terms of voicing, this distinction is primarily phonological. The phonetic distinction, especially in syllable-initial position is more properly characterized as one between voiceless unaspirated and voiceless aspirated stops (Keating, 1984; Lisker and Abramson, 1964; Zlatin, 1974). However, English voicing effects on onset f_0 follow a pattern similar to that found in languages that contrast phonetically *voiced* with phonetically *voiceless unaspirated* consonants. That is, English phonologically voiced (phonetically voiceless unaspirated) stops have the same lowering effect on onset f_0 as do French and Dutch phonologically (and phonetically) voiced stops, while both English phonologically voiceless (phonetically voiceless aspirated) and French and Dutch phonologically voiceless (phonetically voiceless unaspirated) stops have a raising effect on onset f_0 (French: Hombert, 1978; Dutch: Löfqvist *et al.* 1989). This type of patterning (among other factors) led Kingston and Diehl (1994) to argue that consonantal effects on onset f_0 are the result of controlled processes related to the *phonological* status of the consonant series (voiced vs voiceless) rather than a result of intrinsic physiological dependencies between the articulatory and/or aerodynamic properties of the production of different degrees of prevoicing or voicing delay.

Like English, syllable-initial stop consonants in Cantonese may be divided into two phonological voicing classes, typically termed *voiceless unaspirated* and *voiceless aspirated* (Bauer and Benedict, 1997; Lisker and Abramson, 1964). In this case, the phonological and phonetic terminologies agree, and this contrast is distinctive and phonemic. For example, /p/ in the word 錶/piu55/ ("watch") and /pʰ/ in the word 飄/pʰiu55/ ("float") are two different phonemes as these two words carry different lexical meanings (see Sec. I B, for information on tone transcription). In terms of voice onset time (VOT), this Cantonese contrast is similar to the voicing contrast of English. For example, both English and Cantonese voiceless unaspirated stops typically have a short-lag VOT (under 30 ms) and a lower amplitude of the postburst aspiration noise, while voiceless aspirated consonants have a long-lag VOT (greater than 30 ms) and higher amplitude aspiration (Clumeck, Barton, Macken, and Huntington, 1981; Lisker and Abramson, 1964). Based on this similarity,

it seems reasonable to expect that Cantonese, like English, should be considered to manifest the distinction between [+voice] and [−voice] consonant classes in terms of two phonetic features, {voiceless unaspirated} and {voiceless aspirated} (cf. Keating, 1984).

If the effect of consonant voicing on onset f_0 in English is primarily intrinsic, resulting from physiological and/or aerodynamic consequences related to laryngeal gestures involved in voicing (delay), then we would expect Cantonese stops to exhibit the same degree of consonantal influence on onset f_0 as in English because both languages exhibit similar patterns of VOT: Voiced (voiceless unaspirated) stops should show a rising or level onset f_0, and voiceless (voiceless aspirated) stops should show a falling onset f_0. On the other hand, if the English pattern of onset f_0 is (at least partly) due to an intentional manipulation of a secondary cue for the purposes of enhancing the perception of a primary cue (e.g. VOT, cf. Kingston and Diehl, 1994) then Cantonese speakers might show little or no effect of consonant voicing on onset f_0, because f_0 manipulations are likely to be preferentially reserved for tonal cues.

B. Cantonese tones and onset f_0

There are six contrastive tones in Hong Kong Cantonese and each tone has a specific contour (direction of f_0 movement: level, rising, or falling) and register (f_0 height: high, mid, or low) (Bauer and Benedict, 1997; Fok Chan, 1974). While it is not possible to speak of a defined standard for Cantonese, the tonal system in Hong Kong Cantonese differs only slightly from systems observed in speakers from other Cantonese-speaking areas where instrumental phonetic research has been conducted (primarily Macau and mainland Guangdong province). The main difference is the absence of a high falling (51) tone in Hong Kong, where it has been replaced by the high level (55) tone. It appears possible that this is the result of a sound change in progress in Cantonese more generally. Although some speakers from other regions have still been observed to produce the high falling tone consistently, the Hong Kong system is commonly considered the norm (see Bauer and Benedict, 1997 for acoustic data and further discussion). Each Cantonese tone can be represented using the Chao (1947) system with numbers from 1 to 5 expressing the talker's pitch range from lowest to highest. Two numbers are used, such that the first number indicates the pitch level at the start of the syllable while the second number indicates the syllable's ending pitch level. According to this system the tones of Cantonese are as follows. high level (HL): 55; high rising (HR): 25; mid level (ML): 33; low rising (LR): 23; low level (LL): 22; low falling (LF): 21. Given the importance of tone contour in Cantonese, and the large number (four) of tones starting at the same pitch level (2), it is possible that the influence of consonants on onset f_0 may be even more constrained than in other tone languages (e.g., Yoruba, with three phonologically level tones, see Hombert, 1977) because listeners may require more precise information about fine details of a syllable's pitch contour in order to make accurate tonal judgments.

C. Aspiration and onset f_0

Cantonese stops are more properly characterized as differing according to aspiration rather than voicing (see also Bauer and Benedict, 1997). The acoustic properties of the following vowel (f_0, frequency of formant transitions, amount of aspiration noise in the vowel) appear to be more salient perceptual cues than either VOT or (the presence or absence of) aspiration noise for the perception of aspiration in Cantonese initial stops. Tsui and Ciocca (2000) manipulated the duration of the VOT interval of naturally produced aspirated and unaspirated CV syllables to create long VOT conditions with or without aspiration noise between the burst release and the onset of voicing. They found that long VOT stimuli (created by introducing a silent interval between the burst and the onset of voicing of unaspirated stops) were perceived as "unaspirated" by native listeners, showing that VOT per se is a weak cue to the perception of (phonological) aspiration. Poon (2000) systematically manipulated aspiration noise (the level of the breathy sounds between the burst release and the onset of voicing), vowel and VOT cues of similar CV syllables in a fully crossed design, and found that stimuli that contained the vowel portion of syllables originally produced with voiceless aspirated stops were always perceived as "aspirated," independent of the level of the aspiration noise in the gap between the burst release and the onset of voicing and VOT duration. In addition, stimuli that contained a vowel taken from an originally unaspirated syllable were only perceived as "aspirated" if (both) VOT was long and aspiration noise was present between the burst and the onset of voicing, indicating that neither of the two cues (VOT and aspiration noise between the burst and the onset of voicing) could on its own override the effect of vowel type (possibly representing some sort of breathiness property or presence of aspiration noise *during* the first pitch periods of the vowel such as that described for Wu dialects by Jianfen and Maddieson, 1992).

As discussed by Löfqvist *et al.* (1989) and Ohde (1984), gestures associated with consonant voicing contrasts (specifically, increased activity of the cricothyroid muscle for voiceless consonants), not aspiration per se, appear to be the primary physiological basis for consonantal effects on onset f_0. Studies of languages that contrast aspiration, e.g., Hindi (Kagaya and Hirose, 1975) and Korean (Kagaya, 1974), support this hypothesis: Onset f_0 is lower following voiceless aspirated stops as compared with voiceless unaspirated stops, but, at least in Hindi, both types of voiceless stops have a higher onset f_0 than either voiced one. If it is the case that Cantonese stops contrast according to aspiration, not voicing, than we might expect both series of stops to induce some degree of raising of onset f_0.

Even in tonal languages aspirated stops seem to cause less of an increase in onset f_0 than do unaspirated stops. For example, Shi (1998) reported that aspirated stops in Wu and Gan Chinese showed a lower onset f_0 when compared to that of corresponding unaspirated stops, but did not mention the duration of the consonant's effect. Xu and Xu (2003) found that, in Mandarin Chinese, aspirated alveolar stops showed a significantly lower onset f_0 (measured at the first glottal pulse) than did corresponding unaspirated stops, but it is not clear how far into the vowel this effect extended. Finally, Gandour (1974) showed that, in Thai, onset f_0 following voiceless aspirated stops is lower than following voiceless unaspirated stops, and the duration of consonantal effect ranges from 10 to 50 ms. Moreover, in Gandour's (1974) study both voiceless unaspirated and voiceless aspirated stops showed a higher onset f_0 than corresponding voiced stops.

The only study found on the interaction of aspiration and f_0 in Cantonese presented evidence that aspirated stops are associated with a higher onset f_0 than unaspirated stops (Zee, 1980), suggesting that Cantonese does indeed pattern with English with respect to the influence of voicing on onset f_0. However, this groundbreaking study was somewhat limited in scope and investigated only the effects of bilabial consonants on the onset f_0 of syllables with High Level tones. Moreover, Zee (1980) computed only a single value for onset f_0 for each syllable, calculated as the mean value of the first 30 ms after voicing onset. Thus, it is not possible to determine from Zee's (1980) data whether the overall duration of the effect of aspiration was longer or shorter than 30 ms. Such averaging might also have obscured the most significant f_0 changes which were shown by Hombert (1977) to occur immediately after voicing onset.

If Cantonese stops do contrast according to aspiration rather than voicing, then we would expect that both voiceless unaspirated and voiceless aspirated stops should exhibit a relatively high onset f_0; that is, f_0 should fall or at least not rise into the vowel; but unaspirated stops should show a higher onset f_0 than aspirated ones.

D. Goals

The first goal of the present study was to investigate the effect of voicing (or aspiration) differences in Cantonese initial stops on the f_0 of the following vowels, and the interaction of this effect with tone height and contour, and consonant place of articulation. For this purpose, the f_0 of vowels preceded by voiceless aspirated and unaspirated stops was measured over the first 100 ms after voicing onset. Based on the results of previous research, it was hypothesized that there would be a significant difference between the onset f_0 following aspirated and unaspirated stops. Two possibilities of directions of f_0 changes at vowel onset were suggested: (1) a rising versus a falling f_0 pattern would be found to distinguish unaspirated from aspirated stops if the consonantal contrast in Cantonese initial stops is similar to that in English and Yoruba as suggested by Hombert (1977), or (2) the f_0 patterns would be falling for both aspirated and unaspirated stops if the consonantal contrast in Cantonese initial stops is more of an aspiration contrast as in languages including Wu, Gan, Mandarin, Korean, and Thai (Gandour, 1974; Kagaya, 1974; Shi, 1998; Xu and Xu, 2003). In all cases it was expected that any effect of aspiration on f_0 in Cantonese would be shorter than the 100 ms (or so) observed in nontonal languages such as English, French, and Dutch. The goal of the second experiment was to determine whether Cantonese listeners are able to identify consonants in initial

position as either aspirated or unaspirated based solely on differences in onset f_0 and, if so, to estimate the necessary duration of the change in onset f_0 for such a contrast to be made.

II. EXPERIMENT 1: PRODUCTION

A. Methods

1. Subjects

Sixteen native speakers of Cantonese (eight women, eight men) reporting no hearing or speaking disability participated in this study. Participants' ages ranged from 20 to 23 years (mean 21.75). Eight participants (four women, four men) were Hong Kong University students in the Department of Speech and Hearing Sciences with some training in acoustic phonetics. The other eight participants (four women, four men) were students from other departments with no acoustic phonetic training.

2. Stimuli

Ten monosyllabic real words with unaspirated or aspirated initial stops were employed as stimuli. They were: /pa55/ "scar," /pʰa55/ "on all fours," /ta55/ "dozen," /tʰa55/ "he," /ka55/ "home," /kʰa55/ "compartment," /pa25/ "target," /pʰa25/ "steak," /pa21/ "father" and /pʰa21/ "climb." All words had consonant vowel (CV) syllable structure with the vowel /a/ (approximately as in "father" in English) to control for possible effects of intrinsic vowel f_0 (cf. Diehl and Kluender, 1989; Peterson and Barney, 1952). Only real words were used to avoid the difficulty of eliciting nonsense syllables from speakers of a language with lexicographic orthography.

These stimuli can be divided conceptually into two overlapping sets. One set contained /pa55/ and /pʰa55/, /ta55/ and /tʰa55/, and /ka55/ and /kʰa55/. All of the tokens in this set had a high level (55) tone, thus eliminating the possibility of a tone-by-consonant interaction. This set of stimuli was used to investigate the role of place of articulation (bilabial, alveolar, and velar) in determining the effect of consonant aspiration on onset f_0. The second set consisted of /pa55/, /pʰa55/, /pa25/, /pʰa25/, /pa21/, and /pʰa21/. This set of stimuli was designed to investigate the interaction of tone with the effect of aspiration on onset f_0. Due to the constraint of using only real words, it was not possible to generate a complete set of both aspirated and unaspirated CV syllables at a single place of articulation differing according to all six possible Cantonese lexical tones. However, in the present set of stimuli all three tone contours of Cantonese (rising, falling, and level) and both the highest and lowest registers were represented.

3. Procedure

The Chinese character representing each stimulus word was written on a file card with one character per card. Recordings were made with participants seated comfortably in a single-walled sound-shielded booth. Prior to recording, participants read a set of written instructions, and were then familiarized with all of the stimuli used in the experiment. For familiarization, each stimulus card was presented, and

participants were asked to pronounce the target in a carrier phrase with a speaking rate similar to conversational speech. The semantically neutral carrier phrase, /ŋɔ23 wui23 tɔk22 _ _ pɛi35 lɛi23 tʰɛŋ55/ "I will read _ _ for you to hear," was used for both familiarization and recording to encourage a natural production of each stimulus. The results of this study should be expected to generalize to both isolated productions and fluent speech since the absolute f_0 level and contour of syllables in isolation and in carrier phrase context (Ohde, 1984), and in fluent reading (Umeda, 1981) were found to be comparable. Although absolute f_0 values may be affected by coarticulation (Xu, 1994), these effects were held constant since the syllable /tɔk22/ always preceded the target syllable. This preceding syllable has a low level (22) tone, so coarticulatory effects (if present) should cause a uniform lowering of f_0 at the onset of the target syllable across all target syllables (aspirated and unaspirated).

For recording, each of the ten target stimuli was written on five separate file cards for a total of 50 cards (5 repetitions of each of 10 words). All cards were presented to participants individually in randomized order, subject to the constraint that two cards with the same token could not be presented in immediate succession.

Participants were asked to read aloud each word as it was revealed by the experimenter (the third author). Approximately 1–2 production errors were observed for each talker over the course of the 50 target stimuli. After hearing a speech error, the experimenter would immediately produce the target word correctly, and ask the participants to repeat the utterance with that stimulus embedded again. There was no obvious difference between speakers' spontaneously correct productions and those following such prompting. Recording took a total of 15–20 min per talker.

All speech samples were recorded using a Shure Beta 87 microphone and a TASCAM DA-30MKII DAT tape recorder. The microphone was mounted on a boom stand and placed approximately 15 cm in front of the talker's mouth in the horizontal plane. The recorded stimuli were low-passed filtered at 22 kHz and sampled at 44.1 kHz using a Power Macintosh 7200/120AV computer equipped with a DigiDesign Audiomedia II sound card and stored in AIFF format for further analysis.

4. Analysis

Acoustic waveforms, spectrograms, and f_0 plots were obtained using Macquirer 4.9.9 (Macquirer, 1999). For analysis, spectrograms were produced with the Macquirer default wide-band filter setting (172 Hz for male talkers and 344 Hz for female talkers). Fundamental frequency was computed using the default method based on a 35 ms cepstral analysis window. In cases in which the Macquirer algorithm for calculating f_0 was unable to resolve a value at the designated time, f_0 was calculated manually from the acoustic waveform as the reciprocal of the respective glottal period at that time. Plots showing frequency ranges from 50 to 260 Hz for male talkers and 130 to 300 Hz for female talkers were used for analysis. During the measurement of f_0, each recorded stimulus was displayed as an acoustic waveform and a wide-band spectrogram. The onset of voicing was

identified as the first vertical striation (glottal pulse) extending through the first, second and third formants in the spectrogram. This definition of voicing onset is somewhat less reliable, in terms of time, than the onset of periodicity measured from the waveform alone (Francis, Ciocca, and Yu, 2003), but it is a highly conservative marker chosen to maximize the accuracy of automatic pitch-tracking. It should be noted, however, that our reliance on automatic pitch tracking (in most cases), may have resulted in some divergence of our estimate of the duration of f_0 perturbations from that reported by researchers who compute f_0 from direct measurement of single periods (e.g.,Löfqvist *et al.*, 1989; Xu and Xu, 2003). In the case of the present method, a window of 35 ms centered at the point of measurement would result in a margin of error of about 10–15 ms in terms of estimate of the duration of f_0 perturbations. In terms of accuracy of f_0 measurements, our results may be slightly less accurate for female talkers than for males (because we are averaging over more cycles), but interestingly the f_0 patterns look very much the same (see Sec. IC1, below), especially for unaspirated stops. So it is likely that any effect of averaging on precision of f_0 estimate was relatively small.

Following Hombert (1978) the most prominent f_0 changes related to aspiration were expected to be found during the first 100 ms after voicing onset and, because Cantonese is a tonal language, most probably within the first 30–50 ms. Therefore, f_0 was measured at the onset of voicing (time 0) and subsequently at 10, 20, 40, 70, and 100 ms into the vowel. These specific values were selected arbitrarily, but with the goal of maximizing the probability of pinpointing the duration of voicing-related effects on f_0 without requiring an inordinate number of measures.

To permit more accurate comparison across male and female talker groups, prior to statistical analysis frequency values in Hz were converted to cents using the equation $C = 1200 \log_2(f/127.09)$ (Traunmüller, 2005). All analyses of variance involving repeated measures with more than two levels were computed using a standard univariate model with Greenhouse-Geisser and Huynh-Feldt adjustments to degrees of freedom to compensate for the effects of violations of compound symmetry and sphericity (Greenhouse and Geisser, 1959; Huynh and Feldt, 1970). When the degrees of freedom as adjusted by the two methods differed, the more conservative values are reported.

B. Results

Before analyzing f_0, the average overall durations of vowels with each of the three tones in this experiment (HL, HR, and LF) were measured to determine whether vowel duration might interact with tone. Average vowel durations were: HL, 274 ms; HR, 286 ms; and LF, 268 ms, suggesting that f_0 is likely to be a much more important property than duration for identifying these tones. Separate analyses were carried out for the two sets of stimuli: (i) initial stops differing in aspiration and place of articulation, and (ii) initial stops differing in aspiration and in syllables with different tone.
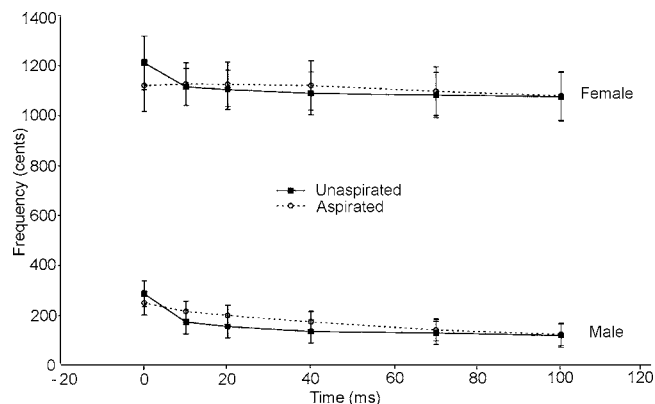
FIG. 1. Mean f_0 for aspirated and unaspirated syllables with high level tone. Error bars indicate standard error of the mean.

1. Aspiration and place of articulation

The first analysis was carried out on stimuli with the same tone (HL) but different places of articulation (bilabial, alveolar, and velar) to determine if place of articulation affects onset f_0. A mixed factorial ANOVA with two levels of aspiration, two levels of gender, three levels of place of articulation (POA), and six levels of time interval showed a significant effect of gender, $F(1,14)=119.59$, $p<0.001$, and of time, $F(1.75,24.46)=50.67$, $p<0.001$, but not of aspiration, $F(1,14)=1.22$, $p=0.29$, or of POA, $F(1.82,25.46)$, $p=0.69$. The only significant interactions were between time and gender, $F(1.75,24.46)=3.60$, $p=0.05$, and aspiration and time, $F(1.47,20.56)$, $p=0.001$. Figure 1 shows the overall average f_0 of tokens with HL tones collapsed over place of articulation during the first 100 ms after voicing onset for the two aspiration categories (aspirated stops $[p^h]$, $[t^h]$,$[k^h]$ versus unaspirated stops $[p]$, $[t]$, $[k]$) for both genders.

Additional posthoc (Tukey HSD) comparisons were performed using means pooled over place of articulation to compare temporally adjacent frequency values (e.g., 0 ms vs. 10 ms) both within and between consonant classes (aspirated vs unaspirated). All results reported as significant are at the $p<0.05$ level. For female talkers, f_0 immediately following unaspirated stops (time=0 ms) was significantly higher (256 Hz) than all other points in both the aspirated and unaspirated series (ranging from 237 Hz to 244 Hz) and this was the only significant pairwise comparison either within or between series. For the male talkers the situation was somewhat more complex. In the unaspirated series, frequency at the first point (time=0 ms) was significantly greater (152 Hz) than all other points in this series, and none of these other unaspirated series' points were significantly different from one another (ranging from 138 Hz to 142 Hz). In the aspirated series no adjacent points were significantly different from one another, but there was a gradual decrease in f_0 such that, for example, the first point (0 ms,148 Hz) was significantly different from the last three (142, 139, and 138 Hz, respectively), and the second (10 ms, 142 Hz) and third (20 ms, 140 Hz) were significantly different from the last two. Comparing the aspirated and unaspirated f_0 curves showed no significant difference between the two series at any measurement point, although a trend similar to that of

the female talkers can be observed: Only at 0 ms was the frequency of the unaspirated series greater than in the aspirated one (151 vs 148 Hz).

Based on the overall similarity of the male and female f_0 contours, the ANOVA was recalculated excluding the between-subjects factor of gender. Results for the combined group were similar to, but less equivocal than, those found in the analysis including gender. There was a significant effect of time, $F(1.58, 23.65) = 43.17$, $p < 0.001$, and a significant interaction between aspiration and time, $F(1.46, 21.83) = 12.34$, $p = 0.001$. No other main effects or interactions were significant. Post hoc (Tukey HSD) analysis showed that the aspirated and unaspirated series differed only at the 0 ms point, with the unaspirated series showing a significantly higher f_0 than the aspirated one. Both curves showed a slight decline in frequency, although this effect is more gradual but also more pronounced in the aspirated series. In the unaspirated series, the first point (0 ms) was significantly higher than all other points, and the second point (10 ms) was significantly higher in frequency than the final two points (70 ms and 100 ms). There were no significant differences found for any pairwise comparison between the final four points (20 ms, 40 ms, 70 ms, and 100 ms) of the unaspirated series, suggesting that the duration of the consonantal effect was no greater than 20 ms. In the aspirated series the first point (0 ms) was significantly higher in frequency than the last three points (40 ms, 70 ms, and 100 ms), while the second (10 ms) and third (20 ms) points were significantly higher in frequency than the final two points (70 ms and 100 ms). The fourth point (40 ms) was significantly higher in frequency than the final point (100 ms), but no other pairwise comparisons among points in the aspirated series showed a significant difference, suggesting that the duration of the consonantal effect was around 40 ms.

In summary, for both male and female talkers, f_0 declined suddenly from 0 to 10 ms in the unaspirated series, but remained more or less level in the aspirated one (declining gradually for the male but not female talkers). For both male and female talkers, frequency patterns across the two series were similar, and generally corresponded to the combined (across gender) analysis in which little or no difference was found between the two series at 10 ms post voicing onset or beyond, but frequency at 0 ms was significantly higher in the unaspirated series than in the aspirated one. Overall, this suggests a pattern of decrease in f_0 following both unaspirated and aspirated stops in Cantonese, but the decline is more abrupt following unaspirated as compared with aspirated consonants.

2. Aspiration and tone

The second analysis was carried out on the measurements of the stimuli with the same place of articulation (bilabial) but different tones (55, 25, and 21). This analysis was designed to investigate whether the differential effects of aspirated and unaspirated consonants on the f_0 of following vowels is realized differently on syllables with different lexical tones. Results are shown in Fig. 2.

A four-way mixed factorial repeated measures ANOVA with factors gender, aspiration, tone (3 levels) and time was
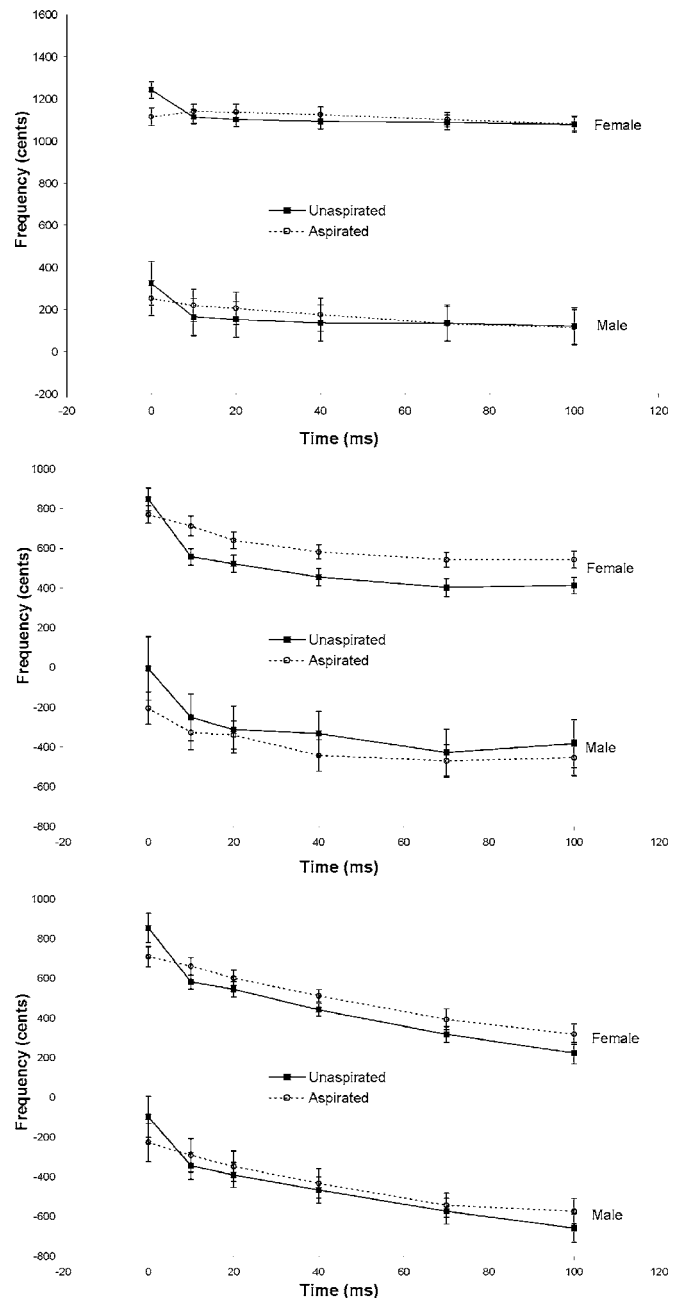
FIG. 2. Mean f_0 for syllables beginning with aspirated and unaspirated bilabial stops and having either high level (2a), high rising (2b), or low level tones (2c). Error bars indicate standard error of the mean. Note that each subgraph shows a Y axis of 1800 cents, though the starting and ending frequencies differ for different tones.

conducted. Results showed a significant effect of gender, $F(1, 14) = 142.39$, $p < 0.001$, tone, $F(1.29, 17.00) = 181.62$, $p < 0.001$, and time, $F(2.07, 28.94) = 152.75$, $p < 0.001$, and significant interactions between tone and time, $F(2.63, 36.78) = 28.52$, $p < 0.001$, and between aspiration and time $F(2.47, 34.54) = 20.77$, $p < 0.001$. No other main effects or interactions were significant.

The observation that gender did not interact with any other variable suggests that both male and female talkers showed the same effects of tone, time and aspiration on onset f_0. Therefore, post hoc (Tukey HSD) analysis were carried out on means pooled over gender. Results showed that, at contemporaneous points across all three tones, the only sig-

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Francis *et al.*: F0 and consonant aspiration   2889

nificant differences due to aspiration were found at 0 ms, where the unaspirated tokens were higher in frequency than the aspirated ones. The only other case in which the aspirated and unaspirated series differed in frequency at the same time was in the 21 (low falling) tone at 100 ms post voicing onset, where there was a significant difference in f_0 between unaspirated (male 87 Hz, female 145 Hz) and aspirated (male 92 Hz, female 153 Hz) consonants.

Results of the post hoc (Tukey HSD) analysis of differences between f_0 measurement points for the high level (55), high rising (25), and low falling (21) tones showed that [results for the high level (55) tone here are for the bilabial tokens only] for the high level (55) tone following unaspirated stops, f_0 was significantly greater at the first (0 ms) measurement point than at all other points, but no other pairwise comparisons showed a significant difference, suggesting that the consonantal effect extends no more than 10 ms into the vowel. For the aspirated series, the first (0 ms) and second (10 ms) points were significantly higher in frequency than the last (100 ms) point, but no other comparisons reached significance suggesting that the effect of the consonant extends no more than 20 ms into the vowel. For the 25 tone, in both series (aspirated and unaspirated) steps 5 (70 ms) and 6 (100 ms) were significantly lower in frequency than the first three steps (0 ms, 10 ms, and 20 ms), but not different from one another or the 40 ms step, suggesting that the consonantal effect ends within 40 ms after voicing onset. For the 21 tone most comparisons reached significance, and in the unaspirated condition even then last step (100 ms) was significantly lower in frequency than the preceding one (70 ms). This result suggests that the consonantal effect may extend up to or beyond 100 ms into the vowel of syllables with a low falling (21) tone (though see discussion, below, regarding the interaction of tones and consonant effects).

C. Discussion

The results of this first experiment are consistent with the hypothesis that consonantal effects on onset f_0 in Cantonese result from a combination of intrinsic physiological/ aerodynamic constraints and of the intentional manipulation of f_0 for the enhancement of voicing cues.

1. Direction of change

All of the syllables investigated in this study showed a generally level or falling f_0 contour over the first 100 ms, consistent with the phonetically voiceless status of both aspirated and unaspirated stops in Cantonese. That is, Cantonese appears to follow the pattern observed in Korean (Kagaya, 1974), Hindi (Kagaya and Hirose, 1975), Thai (Gandour, 1974), and other dialects of Chinese (Wu, Gan: Shi, 1998; Mandarin: Xu and Xu, 2003) rather than that exhibited by English (House and Fairbanks, 1953; Lehiste and Peterson, 1961; Löfqvist et al. 1989): Both long- and short-lag consonants induce a falling f_0 contour at the onset of voicing, though this drop is stronger after unaspirated consonants than after aspirated ones.

These results support two distinct (but probably complementary) physical explanations of the effect of consonant voicing on onset f_0, a muscular/physiological one and an aerodynamic one. According to Halle and Stevens' (1971) "muscular hypothesis," the vocal folds are slack during voiced consonant closures but stiff during voiceless stop closures and subsequent aspiration through the onset of voicing, leading to a higher onset f_0 following voiceless consonants than following voiced stops (see also Löfqvist et al., 1989). Since Cantonese unaspirated stops are associated with a higher onset f_0 than are aspirated ones, Cantonese unaspirated stops are likely to be produced with tenser vocal folds than aspirated consonants, though both may be produced with an overall greater degree of tenseness than, for example, English (phonologically) voiced stops. This conclusion is supported by a number of physiological studies discussed by Shi (1998), suggesting that vocal cord tension at the onset of voicing is higher for voiceless unaspirated stops than for voiceless aspirated stops in Hindi (Kagaya and Hirose, 1975), Thai (Ewan, 1976, cited in Hombert, 1978), and Korean (Kagaya, 1974).

The present findings are also consistent with an aerodynamic account such as that presented by Ladefoged (1971). He argued that the main difference between the production of voiceless aspirated and unaspirated stops is that, at the oral release phase, the glottis is widely opened for aspirated stops, but more or less closed for unaspirated stops. Thus, prior to the onset of voicing more air is released from the lungs through the glottis following aspirated stops than is released following unaspirated stops. Unaspirated stops, which also have a short voice onset time, should exhibit a smaller decrease in subglottal pressure, and therefore a higher initial rate of vocal fold vibration, than aspirated stops prior to the onset of voicing (cf. discussion by Shi, 1998) concerning similar aerodynamic effects in the production of Mandarin initial stop consonants. In the present study, unaspirated stops exhibited a consistently higher onset f_0 than aspirated ones, suggesting that their production does indeed lead to a higher trans-glottal pressure differential at the onset of voicing compared to that of aspirated stops.

Interestingly, while the present results are consistent with a broad range of physiological and aerodynamic accounts of consonantal effects on onset f_0, they appear to conflict with Zee's (1980) findings that Cantonese aspirated stops showed a higher onset f_0 than unaspirated stops. One possible reason for this disagreement is that Zee averaged f_0 over the first 30 ms of each syllable. This averaging process would have made it difficult to identify f_0 patterns similar to those presented here in which the most significant effects were found in the first 10 ms after voicing onset. Indeed, when considering only syllables with a high level tone and beginning with bilabial stops as produced by male talkers in the present study (the conditions most comparable to those of Zee, 1980), averaging the measurements taken at 0, 10, 20, and 40 ms shows an average onset f_0 of 143 Hz for the unaspirated series, and 145 Hz for the aspirated series. While the difference between these values is much smaller than those given by Zee (1980) (average differences of 10, 6, and 12 Hz for three male talkers, based on measurements from

ten productions each of /pei55/ ["sorrow"] and /pʰei55/ ["to spread"]), the trend is similar; moreover, three individuals in the present study exhibited comparable f_0 differences (14 Hz, 5 Hz, and 4 Hz) between aspirated and unaspirated bilabials. Still, the present results strongly suggest that there is little or no physiological contribution of consonant voicing to onset f_0 patterns in Cantonese beyond the first few tens of milliseconds of the vowel.

The findings presented here for Cantonese, and those obtained by Xu and Xu (2003) for Mandarin, also have implications for broader theories of stop consonant production and for the mechanisms of producing a phonological voicing contrast. Keating (1984) proposed a model to resolve the terminological problem that derives from considering, for example, both English and Spanish stop consonants as contrasting according to voicing, despite clear differences in their phonetic realizations. According to Keating's proposal, different languages may instantiate the phonological features [+voice] and [−voice] in different ways such that, for example, a phonologically [+voice] segment in Cantonese or English is realized as a {voiceless unaspirated} stop, while in Spanish a phonologically [+voice] segment would be realized as a {voiced} stop. The results of the present experiment suggest that languages may differ even at the phonetic level, such that the English and Cantonese {voiceless unaspirated} stops have different effects on onset f_0 (lowering in English but raising, if only at the very onset of voicing, in Cantonese). Two possible explanations for this phenomenon have been proposed: Either onset f_0 effects are based on the *phonological* feature involved (such that all [+voice] segments depress onset f_0, while all [−voice] segments raise it), or the Cantonese and English stops may be realized with different patterns of laryngeal muscle movements, each accomplishing laryngeal closure/release in the same time relative to oral closure/release gestures but resulting in different effects on onset f_0.

Previous studies showing an effect of consonant *aspiration* (as opposed to voicing) on onset f_0 have been conducted in languages contrasting more than two phonological voicing classes (e.g., Thai: Gandour, 1974; Hindi: Kagaya and Hirose, 1975; and Korean: Kagaya, 1974); these studies specifically compared phonetically voiceless unaspirated to voiceless aspirated consonants. Interestingly, for at least two of these languages (Hindi and Korean), it has been argued that additional contrastive dimensions beyond that of laryngeal timing must be involved in generating the complete set of contrasts in the language (Abramson, 1977). Kingston and Diehl (1994) distinguish between two voicing-related features, [voice] and [spread glottis], and imply that, e.g. in Thai, the feature [±voice] distinguishes the phonetically voiced ([+voice]) from the phonetically voiceless unaspirated ([−voice]) stops, while the feature [±spread glottis] distinguishes the phonetically voiceless unaspirated ([−spread glottis]) from the phonetically voiceless aspirated ([+spread glottis]) stops. A similar pattern may obtain for Cantonese (and Mandarin), except that Cantonese and Mandarin simply lack the [+voice] category (that is, both the [+spread glottis] and the [−spread glottis] stops are [−voice]). This argument seems to be consistent with the findings of Löfqvist *et al.*

(1989), who showed that the production of both Dutch and English *phonologically voiceless* stops is associated with increased cricothyroid muscle activity. Such an increase in muscle activity presumably increased the longitudinal tension in the vocal folds and resulted in a generally higher onset f_0, although these stops have a short-lag VOT in Dutch (10–15 ms) but a long-lag (60–80 ms) in English. In order to successfully instantiate the long vs short-lag VOT contrast, other laryngeal maneuvers (presumably associated with the feature [±spread glottis]) must be invoked. According to this hypothesis, increased cricothyroid activity is associated with suppression of voicing regardless of the duration of that suppression, and is present in both the voiceless aspirated and voiceless unaspirated consonants in Cantonese, but only in the voiceless aspirated stops in English. The results of the present study suggest that the presence of such cricothyroid activity is automatic, but that the gestures controlling its duration (beyond the first few tens of ms) are not.

2. Duration of effect

In the present study the maximum duration of the effect of consonants on onset f_0 appears to be about 40 ms for the high level (55) and high rising (25) tones, with the effect being somewhat shorter in the high level than in the high rising conditions, while in the low falling condition f_0 continued to decline beyond 70 ms. This pattern is quite commensurate with the results shown by Xu and Xu (2003), and is clearly related to the overall f_0 pattern of each tone. In Cantonese, as in Mandarin, the f_0 contours of level tones typically remain mostly level, sometimes with a slight rise at the onset (Bauer and Benedict, 1997; Xu and Xu, 2003). Thus, these syllables show us the effect of the consonant without any interaction with a tonal contour. In contrast, rising tones typically show a short fall over the course of the first 1/4 to 1/3 of the syllable (Bauer and Benedict, 1997; Francis *et al.*, 2003; Moore and Jongman, 1997; Xu and Xu, 2003), and falling tones typically fall over the course of the entire syllable (Bauer and Benedict, 1997; Xu and Xu, 2003). Thus, in the high level tone, we see the "purest" effect of the consonant, with the f_0 contour more or less leveling off within about 20 ms of the onset of voicing. In the high rising (25) syllables we see the consonantal effect occurring in conjunction with the expected, initially slightly falling tonal contour, leading to the appearance of a consonantal effect out to about 40 ms post voicing onset. Finally, in the low falling (21) tone the longer consonantal effect (70 ms or more in duration) is most likely due to the normal falling f_0 contour for this tone. A close examination of the graphs shown by Xu and Xu (2003) suggests a similar pattern for Mandarin: Consonantal effects do not appear much beyond the first few tens of ms following voicing onset, even in the case of the Mandarin Falling (51) tone.

This finding is consistent with the proposal of Hombert (1978) that tone language speakers have less freedom to vary f_0 patterns as a consequence of voicing (or aspiration). The generally level or falling f_0 pattern in both the aspirated and unaspirated series is consistent with Halle and Stevens' (1971) model of voicing (both series are voiceless). However, the brevity of the effect in Cantonese as compared to

English supports Hombert's (1977) viewpoint that speakers of tone languages actively minimize the intrinsic effect of preceding consonants on f_0 of the following vowels so as to keep each tone maximally perceptually distinct. Alternatively, the aspiration effects on onset f_0 observed in tonal languages might reflect the natural extent of consonant-related influence (extending somewhere between 10 and 50 ms following voicing onset), while the patterns observed for English might result from language-specific exaggeration of these tendencies. On the basis of the present evidence it is not possible to draw any firm conclusions regarding these two conflicting hypotheses, but some speculation is possible.

Onset f_0 differences have been shown to be a sufficient cue to the voicing contrast in English initial stops when the voicing-related f_0 perturbations lasted for about 55 to 60 ms (Haggard *et al.*, 1970). Robinson and Patterson (1995) demonstrated that a minimum of six to eight cycles of a periodic sound are necessary before pitch can be accurately identified. With respect to the present case, this would translate to at least 42–56 ms for a male with an f_0 of 142 Hz, and 25–33 ms for a female with a f_0 of 230 Hz. These durations are consistent with the results presented by Haggard *et al.* (1970), but they are considerably longer than those observed in the present study. If the present results reflect the natural (intrinsic) extent of aspiration effects on onset f_0, then the English pattern (extending out to 100 ms or more) would result from an extension of that natural tendency to nearly twice what would be necessary for listeners to hear the f_0 effects. On the other hand, if the English pattern reflects the natural (unsuppressed) extent of aspiration effects on onset f_0, then the pattern observed here would indicate a much greater degree of suppression than would be necessary to eliminate perceptual evidence of the consonant effects on f_0. A reasonable conclusion then would seem to be that the English pattern reflects some degree of enhancement, while the pattern observed here reflects at least some degree of suppression. In the next experiment, we explore Cantonese listeners' perception of onset f_0 as a cue to the aspiration contrast in syllable-initial stop consonants.

III. EXPERIMENT 2: PERCEPTION

Following Kingston and Diehl (1995), VOT and onset f_0 may "cohere" perceptually (that is, they may both function as cues to the same phonetic feature contrast) either because they arise from the same physiological phenomena in production (cf. Abramson, 1977) or because they affect the auditory system in similar ways (cf. Kingston and Macmillan, 1995). In both cases, the coherence of the two cues may be considered automatic in that it derives from basic physiological properties and does not have to be learned through of linguistic experience. By contrast, it is possible that two cues may cohere due to learned co-variation. That is, over the course of development, listeners may become so accustomed to hearing the two together that the presence of either cue is alone sufficient for perceiving a given contrast, even though there may be no necessary physiological or perceptual relationship between them (cf. Ohala, 1981 and related discussion by Kingston and Macmillan, 1995). In the case of Can-

tonese listeners, who have experienced relatively little perceptible relationship between onset f_0 and VOT in their native language, the ability to use onset f_0 as a cue to the voicing contrast would suggest a basic perceptual (not learned) basis for the relationship between onset f_0 and other stronger cues to aspiration. To investigate this possibility, in the second experiment we investigated Cantonese listeners' ability to use onset f_0 (alone) as a cue to the perception of the Cantonese stop consonant voicing contrast.

A. Methods

1. Subjects

Eighteen native Cantonese speakers, six men and twelve women, reporting normal speech and hearing abilities volunteered for this study. Their age ranged from 22 to 31 (mean age=23.3). All had some experience listening to English in the course of their education (comparable to the experience of the participants in experiment 1), though none used English significantly outside of the university.

2. Stimuli and apparatus

All stimuli were derived from a single CV syllable created by digitally splicing naturally-produced, but resynthesized, burst and aspiration noise onto a synthetic vowel. The natural token (the word /pʰa55/ "on all fours") was selected from a set of many CV syllables produced by a college-aged male native speaker of Cantonese. An aspirated token was selected as the base rather than an unaspirated one to simplify resynthesis. For example, in our experience, it is generally easier (results in a more natural sounding resynthesis) to reduce the amplitude of aspiration noise that is already present than it is to add in aspiration noise that is not there to begin with. A syllable with a high level tone was selected because experiment 1 showed that syllables with this tone do not exhibit an effect of consonant aspiration on onset f_0 beyond the first 10 ms of the vowel at any place of articulation. Thus, Cantonese speakers likely have no prior experience with the covariation of onset f_0 (beyond the first 10 ms) and aspiration in syllables with this tone. The specific token was selected based on the cleanliness of the recording, and the relative smoothness of the acoustic signal (lack of any unexplained acoustic events that might interfere with resynthesis).

The burst and aspiration noise from this token were subsequently manipulated in a manner designed to give rise to approximately equal numbers of /pa55/ and /pʰa55/ percepts (the base syllable). Resynthesis was accomplished using the pitch synchronous overlap and add (PSOLA) algorithm (cf. Huang, Acero, and Hon, 2001, pp. 820–823; Moulines and Charpentier, 1990) as implemented in Praat 4.0.5. To make the burst and aspiration-related cues of the base syllable ambiguous, VOT was compressed to 20 ms and the peak amplitude of the burst and aspiration noise was reduced by 50%, corresponding to values that were shown in an unpublished study in our lab to result in approximately equal numbers of aspirated and unaspirated responses. Informal testing of the final syllable (including the synthetic vowel described below) suggested that the resulting syllable was heard approximately equally as a /p/ or a /pʰ/. The base syllable was
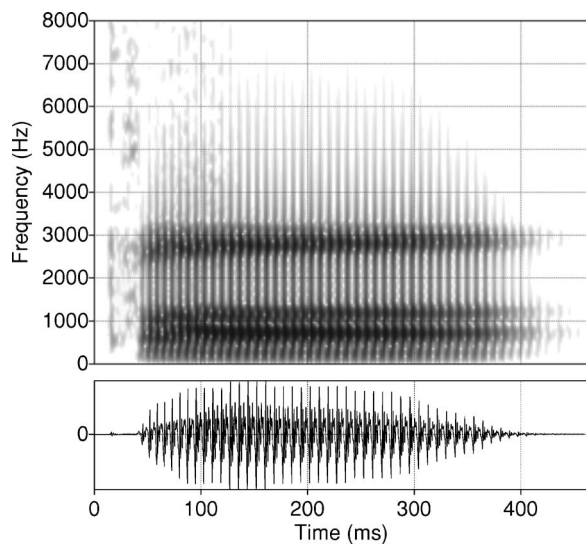
FIG. 3. Spectrogram and acoustic waveform of the syllable [pʰa] with 157 Hz onset f_0 and 80 ms duration of f_0 change, as used in experiment 2.

FIG. 4. Mean proportion of "aspirated" responses to stimuli with f_0 beginning at one of four onset frequencies and one of four durations of onset transition. Error bars indicate standard error of the mean.

420 ms in duration (equivalent to that of the natural syllable), and it was generated using SenSyn (Sensimetrics, Inc.). Synthesis parameters were derived from measurements made at 5 ms intervals of the f_0, amplitude envelope and the first four formant frequencies of the same syllable [pʰa55] "on all fours" from which the burst release was extracted. A spectrogram and acoustic waveform are shown in Fig. 3.

Sixteen different syllables were generated from this base syllable by fully crossing four levels of starting f_0 (127 Hz, 137 Hz, 147 Hz, and 157 Hz) and four levels of f_0 transition duration (10 ms, 20 ms, 40 ms, and 80 ms). For example, in the first stimulus, f_0 began at 127 Hz and fell to the original vowel f_0 (125 Hz) over the course of the first 10 ms following the onset of voicing (ending at a point corresponding to the 30 ms synthesis entry in the synthesizer parameters). The f_0 of the next token also began at 127 Hz, but took 20 ms to reach 122 Hz f_0 (corresponding to an endpoint at the 40 ms point in the synthesized sound). The choice of 157 Hz as the maximum onset f_0 value was based on the mean score of male subjects producing [p] in experiment 1, and subsequent values were selected in arbitrary 10 ms steps below that, encompassing the average range of productions of [p] and [pʰ] by male speakers in experiment 1. The average f_0 of the talker selected for stimulus production in experiment 2 was comparable to the average range of male talkers in experiment 1. All properties of the burst release, aspiration noise, and all properties of the vowel other than f_0 remained the same across all tokens.

3. Procedure

The experiment was carried out in an IAC single-walled sound booth. Stimuli were presented binaurally in randomized order via Sennheiser HD-590 headphones. Stimulus presentation and response collection was controlled by a Hypercard stack running on Macintosh 7100 computer. On each trial listeners heard a single syllable and were instructed to identify it as either /pʰa55/ "on all fours" or /pa55/ "father" by clicking on one of two on-screen button inscribed with the appropriate Chinese characters. Characters were always
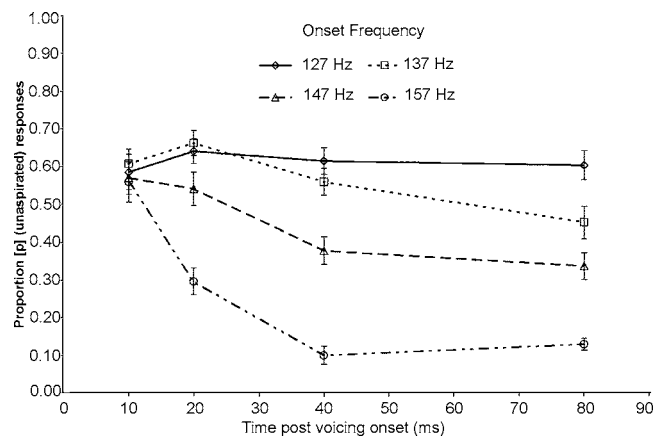
presented side-by-side, in the same order for each participant. Participants completed a total of 16 blocks of 16 trials each (4 levels of onset f_0 times 4 levels of transition duration), with stimulus order randomized within blocks. The first block was treated as a practice block and not scored, although participants were not aware of this at the time of testing. Thus there were a total of 15 scored responses to each stimulus by each participant.

B. Results

Responses were collected and scored in terms of the proportion of "unaspirated" /pa55/ "father" responses for each stimulus, and were submitted to a two-way analysis of variance with repeated measures. Results (shown in Fig. 4) showed a significant effect of onset f_0, $F(3,51)=44.23$, $p<0.001$ and of duration, $F(3,51)=30.42$, $p<0.001$, and a significant interaction between the two, $F(9,153)=11.89$, $p<0.001$. Posthoc (Tukey HSD, $\alpha=0.05$) analysis showed that there was no significant difference in response proportion between any of the tokens with different onset frequencies when the duration of f_0 transition was only 10 ms in duration. For all stimuli with longer f_0 transition durations, the contour with a 157 Hz starting frequency was identified as /pʰa/ "on all fours" (aspirated) significantly more often than any other contour regardless of the duration of the f_0 transition. Listeners identified the 147 Hz contour as /pʰa/ significantly more often than the 127 Hz or 137 Hz contour when the transition was at least 40 ms (though the difference between the 137 Hz and 147 Hz contours disappeared at 80 ms of transition duration). Only at 80 ms transition did listeners perceive the 137 Hz contour as significantly more aspirated (less unaspirated) than the 127 Hz contour.

C. Discussion

Cantonese listeners were able to use onset f_0 as a cue to consonant aspiration in the absence of other, more typical cues. However, there are two important qualifications to this conclusion. First, listeners in experiment 2 were unable to use onset f_0 as a cue to aspiration when onset f_0 varied only

over the 10-ms window shown in experiment 1 to differ between aspirated and unaspirated productions. Thus, Cantonese listeners do not appear to be able to perceive aspiration-related onset f_0 differences that occur in the production of their native language. Second, the manner in which onset f_0 functioned as a cue to consonant aspiration was precisely opposite to what one might have expected if Cantonese listeners' decisions were based on their prior experience with Cantonese. As shown in experiment 1, the onset f_0 of Cantonese unaspirated consonants starts *higher* than that of aspirated consonants, but in experiment 2 stimuli with higher onset f_0 values (e.g., 157 Hz) were generally identified *less* often as unaspirated stops than were those with lower onset f_0 values. Thus, Cantonese listeners' use of onset f_0 in making aspiration decisions must result from something other than their native language experience.

This pattern of results supports a model of voicing/aspiration perception in which onset f_0 can serve as a cue, not because it habitually occurs in conjunction with the cues to the primary contrast (as it does in English), but because it contributes indirectly to some more general, language-independent perceptual phenomenon (cf. Kingston and Diehl's, 1994 "low frequency property"). That is, onset f_0 differences may contribute indirectly to some more general property of the signal that Cantonese listeners are still able to interpret as cuing an aspiration contrast. One factor that may have helped in this transfer or extension of a more general cue is the listeners' experience with English. All of the participants in the present experiment were familiar with English, and used it on a daily basis in the course of their studies at the University of Hong Kong, where they were tested. An attempt was made to minimize the influence of English in the testing context, as all experiments were run in Cantonese and responses were made by clicking on Cantonese characters. Moreover, the English that listeners are exposed to in Hong Kong is typically a non-native variety; most secondary school teachers are native speakers of Cantonese, not English, and the same is increasingly true at the University level. Thus, it is not clear whether the English to which these listeners were exposed even presented a truly English-like pattern of covariation between VOT and onset f_0 cues. However, it is still possible that these listeners obtained some degree of experience with using the low frequency property to distinguish between English voiced and voiceless stops, and that this experience may have contributed to their ability to apply the same cue to the native aspiration contrast in the context of the present experiment.

IV. GENERAL DISCUSSION

The present results provide further support for the hypothesis that the effect of consonant voicing on onset f_0 in English results in part from what Kingston and Diehl (1994) call a controlled process—one that is intentionally manipulated by talkers to produce a desired perceptual effect, and in part from a intrinsic effect of the laryngeal gestures involved in voicing. The intrinsic effect may involve activity of the cricothyroid muscle, perhaps as part of a system for suppressing voicing (Löfqvist *et al.*, 1989), that results in an

increased onset f_0 following voiceless consonants. The active control, in turn, appears primarily to limit the duration of this effect. In the case of voicing in nontonal languages, the role of this active control may be to prolong this effect in order to enhance of consonantal distinctiveness: Presumably, stops tend to sound more voiced when f_0 can be heard to clearly rise into the vowel, and more voiceless when they are clearly followed by a falling f_0. In Cantonese, Mandarin, Thai and other tonal languages, talkers are apparently able to restrict the production of consonant-related perturbations of onset f_0 to the first few tens of milliseconds following the onset of voicing, presumably in order to maintain the integrity of the f_0 contour as a cue to lexical tone identity. The fact that Cantonese listeners are still able to use onset f_0 as a cue for consonantal aspiration differences, despite the lack of such distinctions in their native language, is consistent with Kingston and Diehl's (1994) suggestion that (i) distinguishing voicing/aspiration contrasts may involve attention to a "low frequency property," and that (ii) this property is perceptually enhanced by a variety of independent acoustic cues, including onset f_0. The present results demonstrate that this property (or something similar) is accessible to Cantonese listeners despite their lack of consistent native language experience with onset f_0 differences beyond the first few tens of milliseconds into the vowel.

Since onset f_0 variation beyond the first few tens of ms can be controlled by speakers, longer-term changes in f_0 (e.g., in English) cannot be entirely the involuntary consequences of voicing gestures (i.e., they are not purely intrinsic, physiologically determined, by-products of such gestures). The observation that Cantonese listeners were able to use onset f_0 as a cue to a their native aspiration contrast is consistent with the hypothesis that the use of onset f_0 as a cue to voicing in English and other nontonal languages may derive from some kind of auditory enhancement, perhaps similar to the low frequency property discussed by Kingston and Diehl (1994). However, the present results can also be explained in terms of transfer from listeners' experience with the English pattern of consonantal influence on onset f_0. Future research will be necessary to investigate the use of this cue by native speakers of Cantonese (or other tonal language) who are not familiar with English (or any other language with an English-like pattern of consonantal influence on onset f_0).

Abramson, A. S. (**1977**). "Laryngeal timing in consonant distinctions," Phonetica **34**, 295–303.

Abramson, A. S., and Lisker, L. (**1985**). "Relative power of cues: F0 shift versus voice timing," in *Phonetic Linquistics*, edited by V. Fromkin (Academic, New York).

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin).

Chao, Y. R. (**1947**). *Cantonese Primer* (Harvard University Press, Cambridge).

Clumeck, H., Barton, D., Macken, M. A., and Huntington, D. A. (**1981**). "The aspiration contrast in Cantonese word-initial stops: Data from children and adults," J. Chin. Linguist. **9**, 210–225.

Diehl, R. L., and Kluender, K. R. (**1989**). "On the objects of speech perception," Ecological Psychol. 1, 121–144.

Fok Chan, Y. Y. (**1974**). *A perceptual study of tones in Cantonese*, University of Hong Kong, Centre of Asian Studies, Hong Kong.

Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Francis, A. L., Ciocca, V., and, Yu, J. M. C. (**2003**). "Accuracy and variability of acoustic measures of voicing onset," J. Acoust. Soc. Am. **113**(2), 1025–1110.

Francis, A. L., Ciocca, V. C., and Ng, B. K. C. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**(6), 1029–1044.

Gandour, J. (**1974**). "Consonant types and tone in Siamese," J. Phonetics **2**, 337–350.

Greenhouse, S. W., and Geisser, S. (**1959**). "On methods in the analysis of profile data," Psychometrika **24**, 95–112.

Haggard, M., Ambler, S., and Callow, M. (**1970**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.

Halle, M., and Stevens, K. N. (**1971**). "A note on laryngeal features," Quarterly Progress Report, MIT Research Laboratory, Vol. **101**, pp. 198–213.

Hombert, J. M. (**1977**). "Consonant types, vowel height and tone in Yoruba," Studies in African Linguistics **8**, 173–190.

Hombert, J. M. (**1979**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 77–111.

House, A. S., and Fairbanks, G. (**1953**). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**, 105–113.

Huang, X., Acero, A., and Hon, H.-W. (**2001**). *Spoken Language Processing*, Prentice-Hall, Upper Saddle River.

Huynh, H., and Feldt, L. S. (**1970**). "Conditions under which mean square ratios in repeated measures designs have exact *F*-distributions," J. Am. Stat. Assoc. **65**, 1582–1589.

Jianfen, C., and Maddieson, I. (**1992**). "An exploration of phonation types in Wu dialects of Chinese," J. Phonetics **20**, 77–92.

Kagaya, R. (**1974**). "A fiberoptic and acoustic study of the Korean stops, affricates and fricatives," J. Phonetics **2**, 161–180.

Kagaya, R., and Hirose, H. (**1975**). "Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants," Annual Bulletin, Research Institute of Logopedics and Phoniatrics **9**, 27–46. Available at http://www.umin.ac.jp/memorial/rilp-tokyo/, last accessed May 19, 2006.

Keating, P. A. (**1984**). "Phonetic and phonological representation of stop consonant voicing," Language **60**(2), 286–319.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–494.

Kingston, J., and Diehl, R. L. (**1995**). "Intermediate properties in the perception of distinctive feature values," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 7–27.

Kingston, J., and Macmillan, N. A. (**1995**). "Integrality of nasalization and F1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm," J. Acoust. Soc. Am. **97**, 1261–1285.

Ladefoged, P. (**1971**). *Preliminaries to Linguistic Phonetics* (Chicago University Press, Chicago).

Lehiste, I., and Peterson, G. E. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**, 419–425.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Löfqvist, A., Baer, T., McGarr, N. S., and Seider Story, R. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**(3), 1314–1321.

Macquirer (Version 4.9.9). (**1999**). (Computer software), Scicon R & D Inc., Los Angeles, CA, (www.sciconrd.com).

Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of Mandarin Chinese tones," J. Acoust. Soc. Am. **102**, 1864–1877.

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**(5), 453–467.

Ohala, J. J. (**1978**). "Production of tone", *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 5–39.

Ohala, J. J. (**1981**). "The listener as a source of sound change," in *Papers from the Parasession on Language and Behavior*, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago Linguistic Society, Chicago), pp. 178–203.

Ohde, R. N. (**1984**). "Fundamental frequency as an acoustic correlate of stop consonant voicing," J. Acoust. Soc. Am. **75**, 224–230.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Poon, M. M. W. (**2000**). "Acoustic cues for the perception of aspiration in Cantonese initial stops," B.Sc. dissertation, University of Hong Kong, Hong Kong (unpublished).

Robinson, K., and Patterson, R. D. (**1995**). "The stimulus duration required to identify vowels, their octave, and their pitch chroma," J. Acoust. Soc. Am. **98**(4), 1858–1865.

Shi, F. (**1998**). "The influence of aspiration on tones," J. Chin. Linguist. **26**, 126–145 (in Chinese).

Traunmüller, H. (**2005**). "Auditory scales of frequency representation," http://www.ling.su.se/staff/hartmut/bark.htm. Last accessed June 5, 2006.

Tsui, I. Y. H., and Ciocca, V. (**2000**). "The perception of aspiration and place of articulation of Cantonese initial stops by normal and sensorineural hearing-impaired listeners," Int. J. Lang Commun. Disord. **35**, 507–525.

Umeda, N. (**1981**). "Influence of segmental factors on fundamental frequency in fluent speech," J. Acoust. Soc. Am. **70**, 350–355.

Whalen, D. H. (**1990**). "Coarticulation is largely planned," J. Phonetics **18**, 3–35.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1990**). "Gradient effects of fundamental frequency on stop consonant voicing judgments," Phonetica **47**, 36–49.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "F0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.

Xu, Y. (**1994**). "Production and perception of coarticulated tones," J. Acoust. Soc. Am. **95**(4), 2240–2253.

Xu, C. X., and Xu, Y. (**2003**). "Effects of consonant aspiration on Mandarin tones," J. Int. Phonetic Assoc. **33**, 165–181.

Zee, E. (**1980**). "The effect of aspiration on the fundamental frequency of the following vowel in Cantonese," UCLA Working Papers in Phonetics, Vol. **49**, pp. 90–97.

Zlatin, M. A. (**1974**). "Voicing contrast: Perceptual and productive voice onset time characteristics of adults," J. Acoust. Soc. Am. **56**(3), 981–994.

On first rahmonic amplitude in the analysis of synthesized aperiodic voice signals

Peter J. Murphy

Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland

Rahmonics comprise the prominent peaks in the cepstrum of voiced speech; their locations correspond to the fundamental period and its multiples. The amplitude of the first rahmonic, R1, has previously been used to indicate voice quality. Although a correspondence between R1 and the richness of the harmonic spectrum for voiced speech is well recognized, a formal description has remained absent. A theoretical description of rahmonic analysis of voiced speech containing aspiration noise is provided, leading to a characterization of R1. The theory suggests that R1 is directly proportional to the geometric mean harmonics-to-noise ratio (gmHNR), where the gmHNR is defined as the mean of the individual spectral (i.e. at specific frequency locations) harmonics-to-noise ratios in dB. This hypothesis is validated using synthetically generated voice signals. R1 is shown to be directly proportional to gmHNR (measured directly from the dB spectrum). It is shown that R1 (estimated from speech) is directly proportional to R1 taken from the glottal signal. R1 and gmHNR (measured spectrally) underestimate the actual gmHNR when (averaged) noise levels exceed harmonic levels. Limiting the number of harmonics in the analysis window overcomes this problem and also alleviates the (temporal) window length/f0 dependence of R1 when estimated period synchronously. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2355483]

I. INTRODUCTION

A. Acoustic measures of voice signal regularity

The measurement of regularity of the acoustic voice signal in sustained phonations can provide useful markers for differentiating between normal and pathological voice types (e.g., Kasuya *et al.*, 1986) or for describing general aspects of voice quality (e.g., Hanson, 1997). Temporal regularity of a voice signal resulting from a sustained phonation can be broadly classified in terms of the dimensions of periodicity and wave shape.

Aperiodicity of the voice signal can result from jitter, shimmer, aspiration noise, nonlinear phenomena, interperiod glottal wave shape differences, nonstationarity of the vocal tract, or some combination of these attributes. Measures of these attributes are commonly used to assess pathological voice types. A popular measurement for extracting the degree of aperiodicity of the voice signal is the harmonics-to-noise ratio (HNR); it quantifies the ratio between the periodic and aperiodic components in the signal (Murphy, 1999).

Although a detailed description of voice quality in connected speech entails tracking many time varying acoustic parameters, general acoustic descriptors in the context of sustained phonation have proven useful (cf. Ladefoged and Antonanzas-Barroso, 1985; Klatt and Klatt, 1990; Childers and Lee, 1991; Alku and Vilkman, 1996, Hanson, 1997). These descriptions include temporal and spectral measures that characterize the broad temporal wave shape features of the signal. Measures include H1-H2 [the amplitude difference between the first two harmonics (expressed in dB)], various spectral tilt measures and temporal details of the

glottal flow. These measures, together with aperiodicity measures provide for a broad classification of voice quality.

A number of studies have shown that the amplitude of the first rahmonic peak in the cepstrum (R1) provides a good global descriptor of voice quality. The purpose of the present work is to provide a formal description of R1 and to validate this with the use of synthesis data. The development shows that R1 is directly proportional to the average of the harmonic to between harmonic ratios in the logarithmic spectrum. Before reviewing the use of R1 for voice quality analysis, a brief introduction to the real cepstrum is provided.

B. Voice quality analysis using first rahmonic amplitude

Voiced speech consists of the convolution between a periodic glottal pulse and the combined impulse response of the vocal tract and lip radiation load. Taking the logarithm of the squared amplitude of the Fourier transform of this signal results in a harmonic power spectrum in dB. The cepstral procedure involves taking the inverse Fourier transform of the logarithmic power spectrum. Because the logarithmic power spectrum consists of equally spaced harmonics, a peak in the inverse Fourier transform of this signal (the cepstrum) occurs at a point corresponding to the fundamental period (cf. Rabiner and Schafer, 1978), Fig. 1. The prominent peak at the fundamental period is termed the first rahmonic. Further peaks are termed 2nd, 3rd, etc. rahmonics.

Investigations into the assessment of (pathological) voice qualities and/or aberrant vocal fold vibration using the amplitude of the first rahmonic, R1 include the work of Koike (1986), Koike and Kohda (1991), Hillenbrand *et al.*
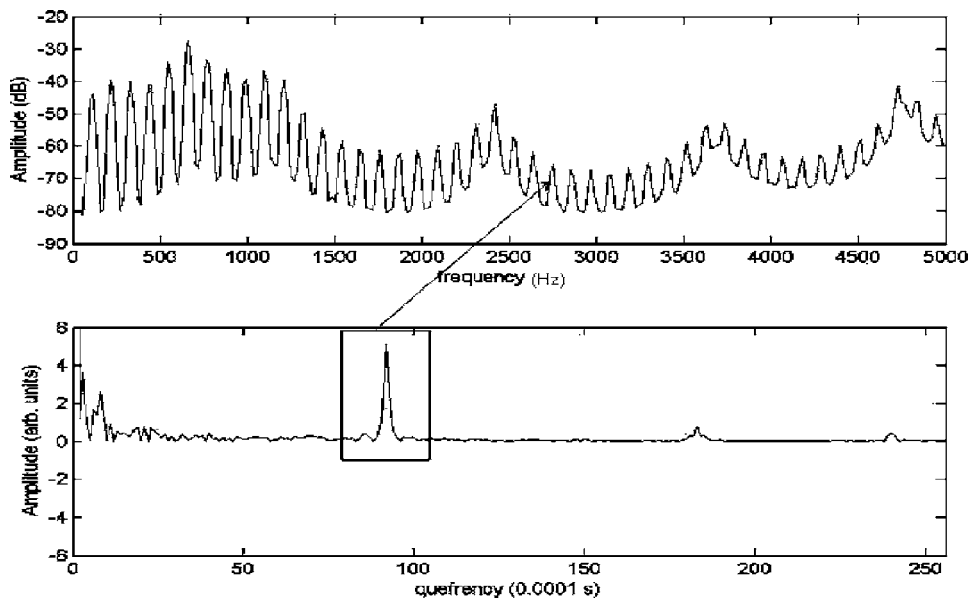
FIG. 1. Spectrum and cepstrum of voiced speech (synthesized vowel /a/ at $f0 = 110$ Hz) illustrating rahmonics at the fundamental period and its multiples.

(1994), Dejonckere and Wieneke (1994), Heman-Ackah *et al.* (2002) and Awan and Roy (2006).[1] Koike and Kohda (1991) observe the effect of surgery to the vocal folds on the prominent cepstral peak in the speech cepstrum. The location of the first rahmonic is used to determine the fundamental period and the height of this peak is used as an indicator of periodicity of the waveform. They state the following:

> "The correspondence between the elongation of the dominant cepstral peak and an improvement of the perceived voice quality also seems noteworthy. Although the regularity of the speech signal can be revealed by the height of the dominant cepstral peak, the scale on the ordinate of the cepstral display is not linear. This makes quantitative analysis of the length of the peaks rather complicated. Quantitative analysis of the height of the cepstral peak has not been attempted in the present study for this reason," p. 264

They note the convenience with which the fundamental period can be measured using cepstral analysis and that the analysis can be performed objectively without human judgment. Dejonckere and Wieneke (1994) note that the amplitude of the harmonic peaks above the noise level in the power spectrum reduce with increasing noise levels and as a consequence the corresponding cepstral peaks reduce in amplitude. They also report that jitter in the waveform leads to reduced harmonic levels and increased between harmonic levels leading to a reduction in the amplitude of the cepstral peaks. They observe that there is no general agreement on the scaling of cepstral results but since the logarithmic power spectrum is expressed in dB that its inverse Fourier transform (the cepstrum) can also be expressed in dB. They report that the main cepstral peak integrates both high frequency noise and perturbation characteristics of pathological voice and hence the measure shows promise for quantifying improvements in voice quality following phonosurgery.

Hillenbrand *et al.* (1994) calculate the height of the first cepstral peak using a normalization procedure and find the resulting index (termed CPP — cepstral peak prominence) to correlate well (correlation coefficient$=0.84$) with the perceptual impression of breathiness. They note that "a highly periodic signal should show a well defined harmonic structure and, consequently, a more prominent cepstral peak than a less periodic signal," p. 772. Herman-Ackah *et al.* (2002) report that the CPP index provides an indication of the harmonic structure for voiced speech and that it correlates strongly with perceptual measurements of dysphonia. In comparison with alternative acoustic indices (e.g., jitter, shimmer, and various spectral tilt and noise measures) Herman-Ackah *et al.* (2002), Hillenbrand *et al.* (1994), Dejonckere and Wieneke (1994) and Awan and Roy (2006) independently report that the cepstral peak provides a superior correlate of overall voice quality.

In the studies cited above that use R1 it is observed that increased levels of noise and perturbation in the voice signal are reflected by a reduction in the measure. However, it is also noted that a formal description of the cepstral peak remains absent. The present study provides a theoretical description of cepstral analysis of voiced speech containing aspiration noise, which leads to a characterization of R1 as a measure directly proportional to a geometric mean harmonics-to-noise ratio (gmHNR). The technique is then tested on a set of synthetically generated voice signals containing various aperiodicities. In a preliminary validation, R1 is correlated against gmHNR measured directly from the dB spectrum. A number of spectral analysis procedures are implemented, including period asynchronous, period synchronous, and a harmonic-limited procedure. A discussion is given in terms of the present results on synthesis data and on the consequence of the characterization on the results of previous studies involving human voice signals.

II. METHOD

Noll (1967) pointed the way forward to a practical interpretation of the cepstrum stating that "the spectrum itself can be regarded as a signal and can be processed by standard signal-analysis techniques" (p. 309). The present description
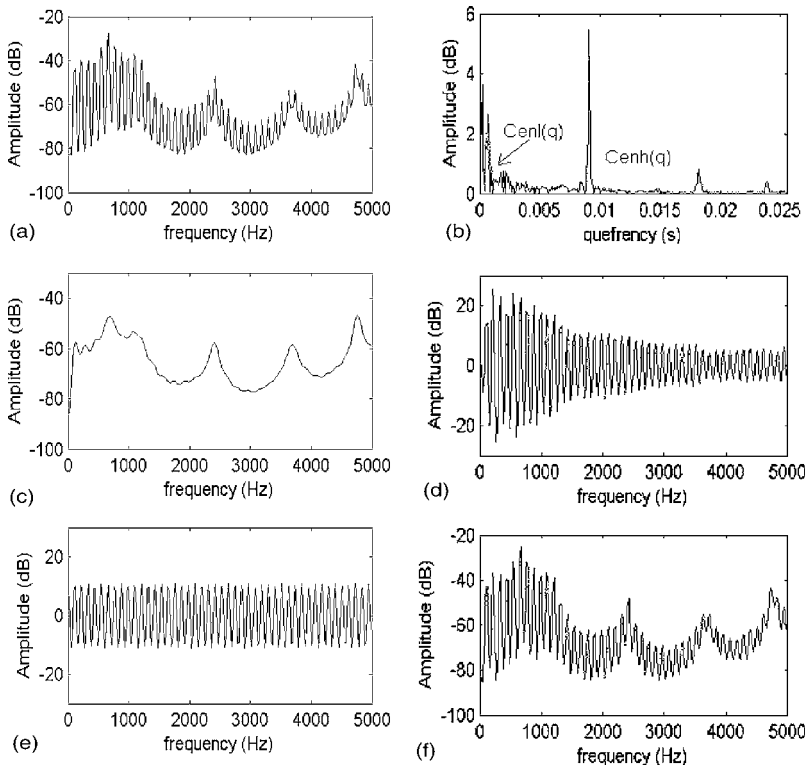
FIG. 2. (a) Averaged spectrum for synthesized vowel /a/ ($f0 = 110$ Hz) containing 1% glottal noise (512 sample points sampled at 10 kHz, 50% overlap, 1.5 s). (b) Cepstrum of the signal in (a). (c) Fourier transform of low-passed liftered cepstrum [$C_{enl}(q)$]. (d) Fourier transform of cepstrum liftered in the region of the first rahmonic. (e) Fourier transform of first rahmonic peak. (f) Sum of Fourier transforms of liftered cepstra in (c) and (d).

considers taking the inverse Fourier transform of the logarithmic power spectrum of voiced speech containing aspiration noise. This development leads to a characterization of R1, which suggests that it is directly proportional to a geometric mean harmonics-to-noise ratio.

A simple representation for voiced speech containing aspiration noise [$s_{en}(t)$] can be written as

$$s_{en}(t) = \{[e(t) * g(t)] + n(t)\} * v(t) * r(t), \tag{1}$$

where $e(t)$ is a periodic impulse train, $g(t)$ is a single glottal pulse, $v(t)$ and $r(t)$ represent the vocal tract impulse response and the radiation load, respectively, $*$ indicates convolution, $s_{en}(t)$ is a periodic voiced speech waveform plus aspiration noise, and $n(t)$ represents aspiration noise.

Applying a window, (w), gives

$$s_{en}^w(t) = \{[e(t) * g(t)] + n(t)\} * v(t) * r(t) \times w(t). \tag{2}$$

The window function, $w(t)$, can be moved inside the convolution, provided the window length is sufficiently long (Oppenheim et al., 1968), to give

$$s_{en}^w(t) = \{[ew(t) * g(t)] + n_w(t)\} * v(t) * r(t). \tag{3}$$

Taking the Fourier transform gives

$$S_{en}^w(f) = [E_w(f) \times G(f) + N_w(f)] \times V(f) \times R(f). \tag{4}$$

where capital letters indicate the Fourier transforms of the corresponding temporal descriptors. Taking the logarithm of the magnitude squared values and approximating the signal energy at harmonic locations, $\log|S_{en}^w|_h^2$ and at between-harmonic locations, $\log|S_{en}^w(f)|_{bh}^2$, gives

$$\log|S_{en}^w(f)|_h^2 = \log|E_w(f) \times G(f) \times V(f) \times R(f)|^2, \tag{5}$$

$$\log|S_{en}^w(f)|_{bh}^2 = \log|N_w(f) \times V(f) \times R(f)|^2. \tag{6}$$

Although noise is present at all frequencies in the spectrum, when spectral averaging is applied, the harmonics approach the true harmonic values and the between-harmonics approach the true noise variance [Fig. 2(a)]. In the spectral averaging process the voiced speech signal is broken up into frames and successive frames are overlapped. A spectrum is produced for each frame and an average of these spectra is taken.

Taking the cepstrum of $s_{en}^w(t)$ gives

$$C_{en}(q) = \text{IDFT}[\log|S_{en}^w(f)|^2], \tag{7}$$

where IDFT is the Inverse Discrete Fourier Transform and "q" indicates quefrency.

The cepstrum translates to low and high quefrency components:

$$C_{en}(q) = \begin{cases} C_{enl}(q), & 0 < q < q_c, \\ C_{enh}(q), & q_c < q < q_{max}, \end{cases} \tag{8}$$

where "q" indicates quefrency, q_c indicates the cutoff point between high and low quefrency regions, and the subscripts "l" and "h" indicate low and high quefrency, respectively (Fig. 2). The low quefrency region is specified from 0 ms up to a quefrency corresponding to the first rahmonic minus the width of its main lobe and significant side lobes.

The low quefrency cepstral coefficients [$C_{enl}(q)$] reflect a combination due to the glottal flow excited vocal tract plus radiation [the last three terms in Eq. (5)] and the noise excited vocal tract plus radiation [Eq. (6)]. The high quefrency [$C_{enh}(q)$] is characterized by the rahmonic peaks spaced at the fundamental period and its multiples [Fig. 2(b)]. This separation of speech components in the cepstrum allows for

various filtering (liftering in the quefrency domain) operations. Figure 2(c) shows the Fourier transform of the cepstrum low-pass liftered below the first rahmonic (samples after 6.2 ms are set to zero) while Fig. 2(d) shows the Fourier transform of the cepstrum that has been set to zero everywhere except in the region (±3 ms) of the first rahmonic. The latter figure provides an approximation to the harmonics-to-noise structure in the original spectrum, with the spectral envelope removed. These Fourier transformed liftered signals [Fig. 2(c) and 2(d)] are added in Fig. 2(f), providing a close approximation to the original spectrum in Fig. 2(a).

The amplitudes of the rahmonics reflect the richness of the harmonic structure (in logarithmic scale). Specifically, the amplitude of the first rahmonic, R1, is dependent (as are all the rahmonics) on the depth of the valleys between adjacent harmonic locations [consider taking the inverse Fourier transform of the spectrum in Fig. 2(a)]. The amplitudes of the harmonic locations are determined by $|E_w(f) \times G(f) \times V(f) \times R(f)|^2$ while the amplitudes of the "between harmonic" locations are determined by $|N_w(f) \times V(f) \times R(f)|^2$. Therefore, the noise contributions can be seen to limit the height of the first rahmonic amplitude. Consequently, estimating the amplitude of the first rahmonic provides some indication of the harmonics-to-noise ratio content of the signal. R1 is determined by the average of the difference between harmonic and "between harmonic" energies in the dB spectrum, and hence it is not directly related to the HNR of the radiated speech waveform because it is, to a first order approximation, independent of the envelope in the original spectrum. Furthermore, because the spectrum is in dB, the resulting index is proportional to an average of dB values. Figure 2(e) shows the Fourier transform of the cepstrum, where every point except the first rahmonic peak has been set to zero. The amplitude of the resulting signal can be seen to be an average of the signal in Fig. 2(d). Finding an average of dB values is equivalent to determining the geometric mean HNR of the power spectrum (gmHNR — expressed in dB). The term geometric mean comes from the fact that the averaging involves logarithms and is therefore equivalent to taking the product of linear values and taking the Mth root (and then taking the logarithm), where M refers to the number of harmonic to between-harmonic estimates.

The geometric mean harmonics-to-noise ratio (gmHNR) is defined as

$$\text{gmHNR} = \frac{20}{M}\sum_{f=0}^{M} \log 10\left(\frac{|E(f) \times G(f) \times V(f) \times R(f)|}{|N(f) \times V(f) \times R(f)|}\right).$$

(9)

This can be approximately written as[2]

$$\text{gmHNR}' = \frac{20}{M}\sum_{f=0}^{M} \log 10\left(\frac{|E(f) \times G(f)|}{|N(f)|}\right),$$

(10)

and, hence, the measure is, to a first approximation, independent of the vocal tract resonance and radiation load, which is desirable for quantifying a glottal source signal-to-noise ratio. The geometric mean harmonics-to-noise ratio, gmHNR,

is equivalent to a conventional HNR estimate of the glottal source (GHNR) only when the HNR is constant for all frequencies. This is satisfied when shimmer (only) is present in the glottal source signal (Murphy, 2000). In general, the measure represents a geometric mean of the linear HNRs determined for each harmonic/between-harmonic of the glottal source (with the overall result being expressed in dB). Equivalently, it represents an arithmetic mean of the dB HNRs for each harmonic/between-harmonic location.

The rahmonic analysis development suggests that R1 is directly proportional to gmHNR, i.e.,

$$\text{R1} \propto \text{gmHNR}. \tag{11}$$

A direct proportionality rather than equality is assigned in Eq. (11) because the spectral wave shape (harmonic - between harmonic pattern) is not a pure sinusoidal signal but rather has a characteristic wave shape depending on the fundamental frequency and the temporal analysis window length.

III. ANALYSIS

A. Synthetic speech signals

A sequence of glottal pulses (type C, Rosenberg, 1971) is used as input into a delay line digital filter, with filter coefficients set to represent the vowel /a/ (Rabiner and Schafer, 1978). Radiation at the lips is modeled by the first order difference equation $R(z) = 1 - z^{-1}$. The sampling frequency is 10 kHz and 1.5 s are synthesized.

Aperiodicity is introduced by altering the glottal source in a 110 Hz waveform. Random shimmer is introduced by adding a random variable gain factor (of a given std. dev. ranging from 1% to 32%) to the amplitude of the fundamental period impulse train prior to convolution with the glottal pulse. Random jitter is synthesized at levels from std. dev. 1% to 6 % through time scaling of glottal waveforms. Random noise is introduced via a random noise generator arranged to give mean zero white Gaussian noise of a user specified variance (std. dev. 1% to 32%). Further signals are also created for three levels of random noise (4%, 8%, and 16%) for signals beginning at 80 Hz and increasing in six, approximately equispaced steps of 60 Hz up to 350 Hz.

B. Spectral and cepstral processing

The synthesis signals are analyzed using period-asynchronous and period-synchronous windowing. For the asynchronous analysis Hamming window lengths of 256, 512, 1024, 2048, and 4096 are used. These frames are hopped by half this window length (e.g., by 256 points for a 512 point window) for the 1.5 s available. For the period-synchronous analysis 4, 6, 8, 10, and 12 fundamental periods are used and hopped by half this number. The default value of fundamental period is used, making the method essentially period asynchronous when the jitter signals are analyzed. The linear spectra resulting from the hopped windowed segments are averaged and a single dB spectrum is calculated. The resulting dB spectra are inverse Fourier transformed to provide the cepstral coefficients and R1 is located using a peak-picking algorithm. The peak-picking algorithm simply

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Peter J. Murphy: Rahmonic analysis of voice   2899
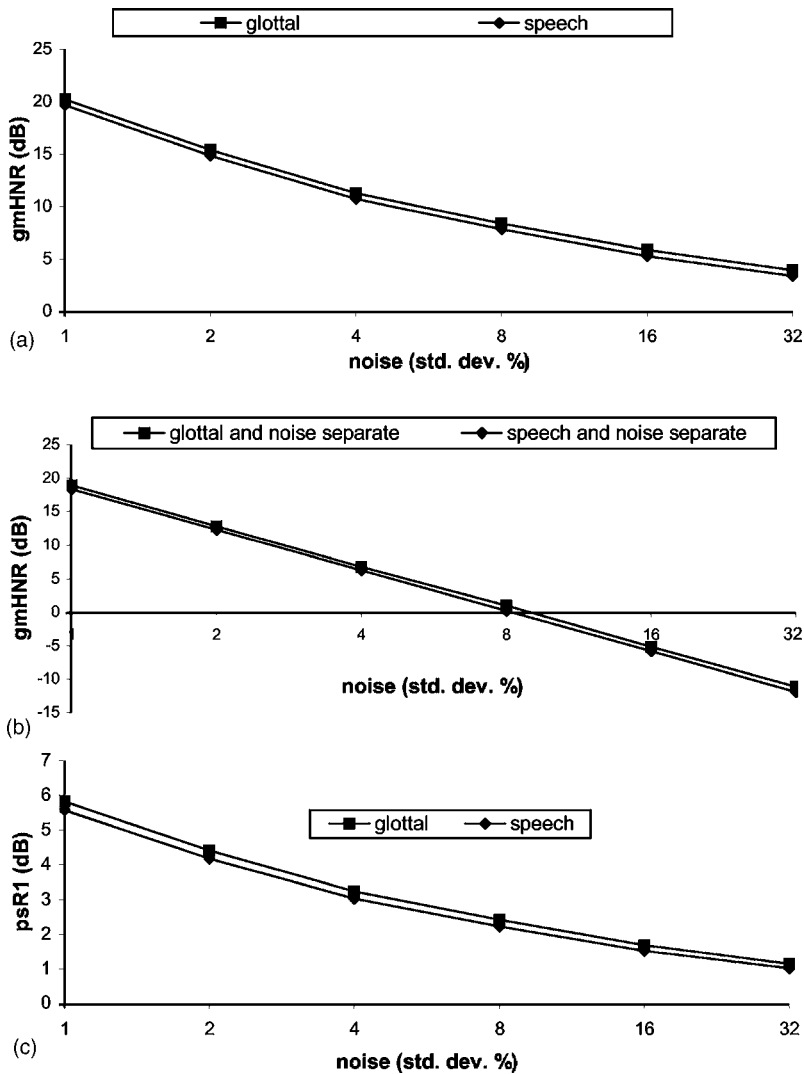
FIG. 3. Period synchronous analysis (six periods) for synthesized glottal pulse and vowel /a/ at 110 Hz for six levels of glottal noise (a) gmHNR (harmonics and noise estimated from noisy signals) (b) gmHNR (harmonics and noise estimated from separate signals: noise-free glottal and speech signals and noise-only signals) (c) psR1 (estimated from noisy signals).

searches for the maximum amplitude in the region of the expected fundamental period, T0, which is known for the synthesis data. The amplitude is taken directly from the cepstrum without further processing. For the dB spectra produced (for a window length of six fundamental periods) a spectral measure of gmHNR is calculated through estimating the harmonic and between-harmonic levels directly.[3]

IV. RESULTS

In order to test the algorithms in a systematic manner, synthetically generated glottal and speech signals are analyzed. The unit for R1 is the dB as it represents a mean of dB values. In order to validate that gmHNR (calculated from the dB speech spectrum) and gmHNR (calculated from the dB glottal spectrum) are approximately equal [Eq. (9) and Eq. (10)] and to validate that R1 is directly proportional to gmHNR, gmHNR is estimated directly from the dB spectrum (i.e., signal energy is estimated at harmonic locations and noise is estimated halfway between harmonic locations). For this analysis "x" % noise is added to the glottal source and the averaged spectrum (expressed in dB) is produced for the corresponding glottal and speech signals (a period synchronous window length of six periods, hopped by three periods, is used; $f0=110$ Hz). gmHNR is extracted from the resulting

spectra as shown in Fig. 3(a). In Fig. 3(b) gmHNR is measured by estimating the harmonic peaks from noise-free spectra [i.e., from the averaged spectra (expressed in dB)] of the glottal and speech signals with no noise added to the glottal source. Similarly, the between-harmonic (i.e., noise) estimates are taken directly from noise only dB spectra (no glottal or speech signals present). A Spearman's correlation of gmHNR (speech) and gmHNR (glottal) gives $\rho=1.00$ significant at the 0.01 level for the data in Fig. 3(b). In Fig. 3(c), psR1 (period synchronous R1) is estimated from the cepstrum using the same spectra as used for gmHNR in Fig. 3(a). A Spearman correlation provides a $\rho=1$, significant at the 0.01 level for all paired correlations between gmHNR (glottal, speech) and psR1 (glottal, speech).

The R1 values resulting from a period-asynchronous cepstral analysis (window lengths 256, 512, 1024, 2048, and 4096) of voiced speech ($f0=110$ Hz) with six levels of random noise (1% in doubling of levels to 32%) are shown in Fig. 4(a). At a given noise level (e.g., 2%), R1 decreases as the window length increases (except for the lowest window size, 256). R1 (period-asynchronous, window length 1024) is shown to be approximately equispaced and somewhat invariant across $f0$ when tested against three noise levels for six different fundamental frequencies [Fig. 4(b)].
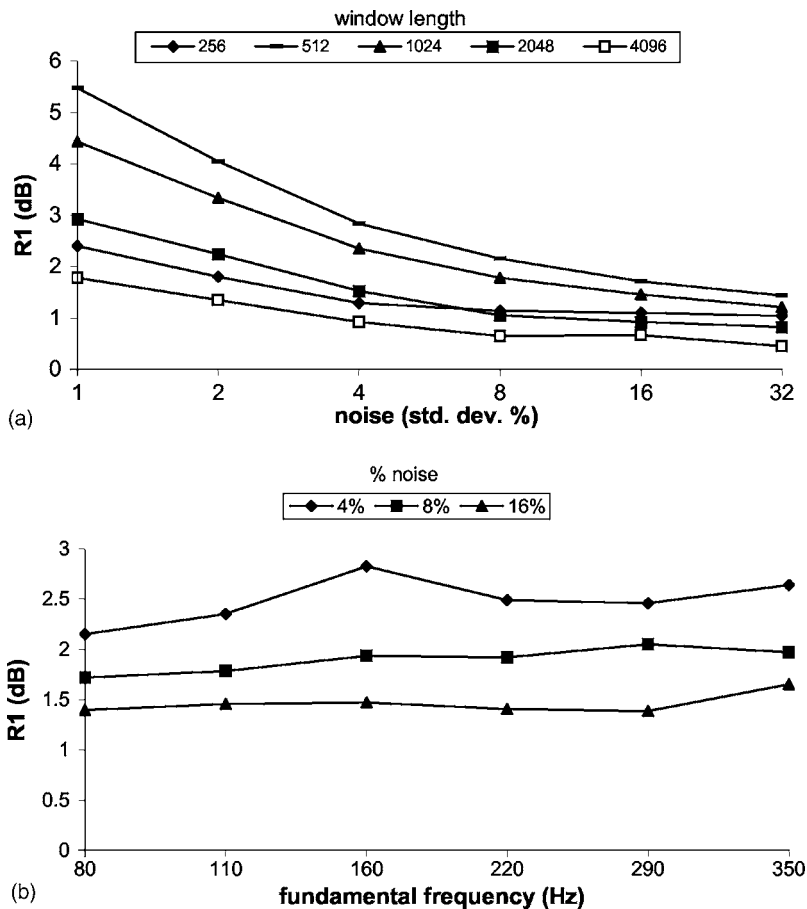
FIG. 4. (a) Period asynchronous R1 for a 110 Hz synthesized vowel /a/ with six levels of glottal noise and five window lengths. The linear spectra resulting from the hopped windowed segments are averaged and a single dB spectrum is calculated from which R1 is estimated. (b) Period asynchronous R1 plotted against $f0$ for the synthesized vowel /a/ with three levels of glottal noise (window length=1024).

Period-synchronous R1 (psR1) is tested for the 110 Hz signal [Fig. 5(a)] for window lengths of 4, 6, 8, 10, and 12 fundamental periods and for the range of $f0$ values using a window length of six fundamental periods [Fig. 5(b)]. In all subsequent period synchronous analyses, a window length of six fundamental periods is used. In Fig. 6 the spectrum is limited to the first 13 harmonics and psR1 is shown to be equispaced and largely invariant across $f0$ for three noise levels for the range of $f0$ values (the full 1.5 s data record is used for each $f0$).

Figure 7 shows a linear response of (a) gmHNR and (b) psR1 for glottal and speech waveforms (spectrum limited to the first 13 harmonics) with respect to the doubling of noise levels. Figure 8 shows that gmHNR and psR1 (for glottal and speech data using complete and harmonic limited spectra) respond linearly with the doubling of shimmer levels. Figure 9 shows that gmHNR and psR1 (for glottal and speech data using complete and harmonic limited spectra) decrease nonlinearly with respect to linear increases in jitter levels. A Spearman correlation gives $\rho=1.00$, significant at the 0.01 level for all paired correlations between gmHNR (speech), psR1 (speech), gmHNR (glottal), and psR1 (glottal) versus noise, shimmer and jitter, respectively, when estimated from complete and harmonic limited spectra.

V. DISCUSSION

A. Results for synthesis data

The key to understanding the results of the cepstral analysis lies in examining the underlying spectral character-

istics. Figure 3(a) validates that gmHNR (measured directly from the dB spectrum) is directly proportional for the speech and glottal signals (they have almost the same values). This is noteworthy because it indicates that an acoustic measurement taken from the voiced speech waveform can potentially be used to infer information directly related to the glottal source, which is desirable from a noninvasive, clinical diagnostic viewpoint. Figure 3(c) similarly shows that psR1 is directly proportional for the speech and glottal signals and a Spearman correlation between gmHNR and psR1 provides a $\rho=1$, significant at the 0.01 level, confirming that the hypothesis put forward under METHOD is correct. However, it is noted that gmHNR and psR1 (in contrast to a conventional HNR estimate) do not vary linearly with the doubling of noise levels. Equation (9) and Eq. (10) suggest that a doubling of noise levels would lead to a linear decrease in the gmHNR value. The nonlinear decrease in psR1 and gmHNR is due to the fact that as noise levels exceed harmonic levels (at higher frequencies) in the averaged spectrum the level of harmonics are overestimated, leading to overestimates of gmHNR.[4] Figure 3(b) illustrates gmHNR extracted from the spectrum where harmonics are estimated from noise free, glottal, and speech signals. The noise is estimated from glottal and speech free signals (i.e., noise only is present). The resulting response of gmHNR with respect to the doubling of noise levels is linear (in this case the ratio decreases in steps of approximately 6 dB/doubling of noise levels, as per a conventional HNR estimator).

In the asynchronous analyses illustrated in Fig. 4, the
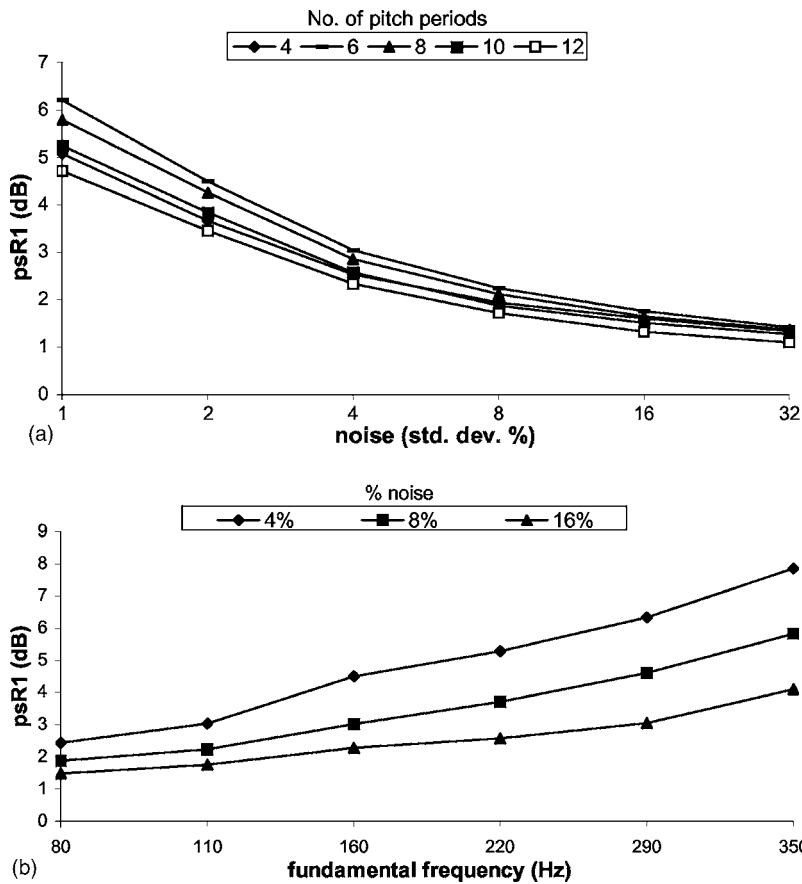
FIG. 5. (a) Period synchronous R1 (psR1) versus glottal noise levels for varying window lengths. (b) Period synchronous R1 (psR1—six periods) versus fundamental frequency for three glottal noise levels.

nonlinear decrease of R1 with respect to the doubling of noise levels is again observed. As the window length increases (ignoring 256 initially), the value of R1, at a given noise level, decreases. This can be interpreted as the increased harmonic resolution leading to a decreased sinusoidal nature of the spectrum [Figs. 10(b) and 10(c)]. Conversely, the window length=256 provides a lower R1 due to a reduction in spectral resolution, i.e., even though the spectrum is somewhat sinusoidal in nature (compared to window length=1024, for example) the troughs are not as steep as for the case of window length=512 due to the decreased spectral resolution (Fig. 10). For the longer window lengths (2048 and 4096) as the noise level increases beyond 8% there is only a slight variation in R1. This happens at 4% for the 256 sample window length.
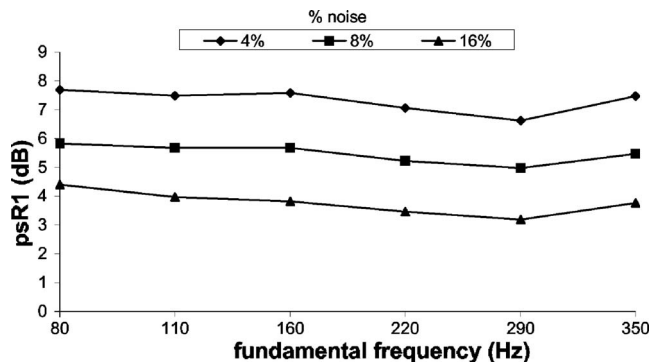
FIG. 6. Period synchronous R1 (psR1) with the spectrum limited to the first 13 harmonics versus fundamental frequency for three levels of noise.

The variation of R1 with respect to $f0$ is examined in Fig. 5 (window length=1024 points). It has been stated previously that gmHNR is, to a first approximation, independent of the vocal tract, hence it might be anticipated that R1 is independent of $f0$. The plot of R1 vs $f0$ for three levels of random noise indicates that the variation is not that of a conventional HNR. The trend of increasing HNR with increasing fundamental frequency for synthetic speech signals has been observed previously and is due to the difference in excitation of the vocal tract formants as the fundamental frequency increases (Murphy, 1999). Although the measurements of R1 are reasonably consistent across $f0$, some variation persists, as indicated in the figure. In fact, the seemingly flat response of R1 vs $f0$ can be seen to result from a counterbalancing of two aspects related to the spacing of the harmonics in the spectrum. As $f0$ increases, R1 potentially increases due to an overall increase in the average harmonics-to-noise ratios in dB over the frequency range of interest (in this case 0 to 5 kHz), whereas it decreases due to the increased frequency resolution that results from the harmonics being spaced farther apart [Figs. 11(a) and 11(b)].

Period synchronous R1 (psR1) is shown for the 110 Hz signal in Fig. 5(a). It shows a similar trend to the asynchronous analysis with psR1 reducing nonlinearly as noise levels double. Except for the case of four fundamental periods (91 samples at 10 kHz sampling×4), as the window length increases R1 decreases. As for the period-asynchronous case, this trend is due to harmonic resolution. Six fundamental periods are chosen for further analysis. The variation of psR1 vs $f0$ is shown in Fig. 5(b). This trend is similar to the
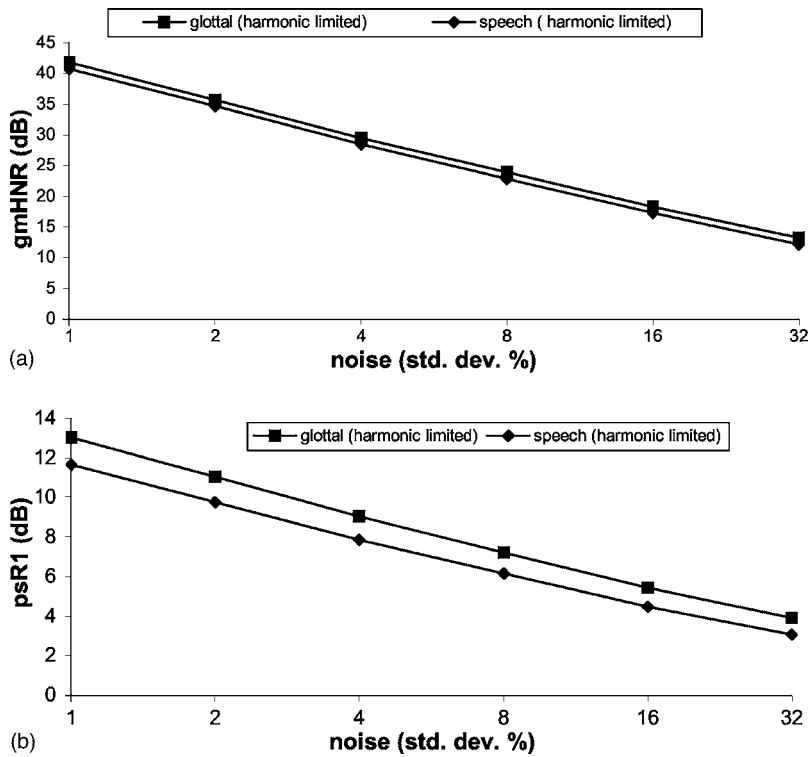
FIG. 7. (a) gmHNR (spectrum limited to the first 13 harmonics) and (b) psR1 (spectrum limited to the first 13 harmonics) versus noise for the 110 Hz synthesized glottal and speech signals.

variation of HNR vs $f0$ (Murphy, 1999), however, the origin of the trend is different. In a conventional HNR measurement, the index is dependent on the vocal tract resonances, whereas the gmHNR is, to a first approximation, not. The $f0$ dependence of the index in this case arises from the spacing of the harmonics over the frequency range of interest [Figs. 11(c) and 11(d)]. From the figure it can be seen that the 110 Hz signal has many smaller dB HNRs than the 220 Hz signal, due to the greater number of harmonics, leading to a lower value of R1. Because the measure is period synchronous in this case the harmonic resolution is equal across $f0$ and hence does not offer the counterbalancing of the asynchronous analysis.

As stated previously, ideally the acoustic analysis should
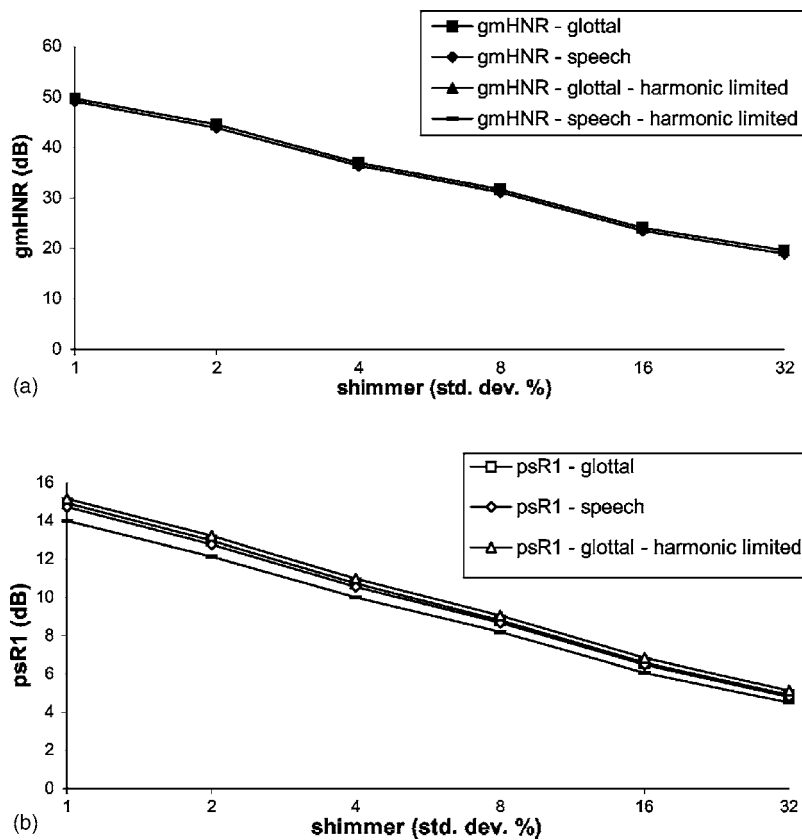
FIG. 8. (a) gmHNR (complete spectrum and spectrum limited to the first 13 harmonics) and (b) psR1 (complete spectrum and spectrum limited to the first 13 harmonics) versus shimmer for the 110 Hz synthesized glottal and speech signals.
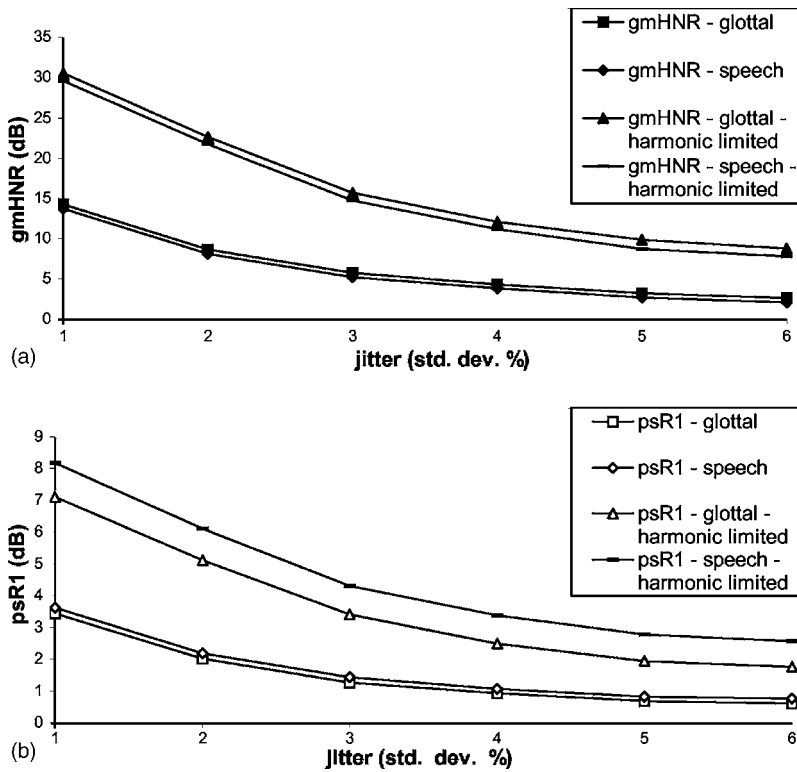
FIG. 9. (a) gmHNR (complete spectrum and spectrum limited to the first 13 harmonics) and (b) psR1 (complete spectrum and spectrum limited to the first 13 harmonics) versus jitter for the 110 Hz synthesized glottal and speech signals.

reflect levels of source aperiodicity, independent of $f0$. In an attempt to make psR1 an $f0$ independent measure, a set number of harmonics ($k=13$) is taken in the spectrum prior to taking the inverse Fourier transform (Fig. 6). An approximately flat response with respect to $f0$ is observed. A second important advantage of limiting the analysis to a set number of harmonics is illustrated in Fig. 7. The graph shows that (a) gmHNR (speech and glottal—spectrum limited to the first 13 harmonics) and (b) psR1 (speech and glottal—spectrum limited to the first 13 harmonics) decrease linearly with the doubling of noise levels. By limiting the spectrum to, in this case, the first 13 harmonics, the harmonic levels are not ex-

ceeded as the noise level increases. The (harmonic limited) gmHNR values are almost the same at a given noise level for the glottal and speech data. However, psR1 estimated from the harmonic limited spectrum for the glottal signal is approximately 1 dB greater than when estimated from the harmonic limited speech spectrum. This is because of the zero Hertz or dc term for the glottal signal. In the estimation of gmHNR, the dc term is ignored. The dc term can be included in gmHNR or, conversely, the harmonic spectrum could be band limited to begin at the first harmonic prior to taking the
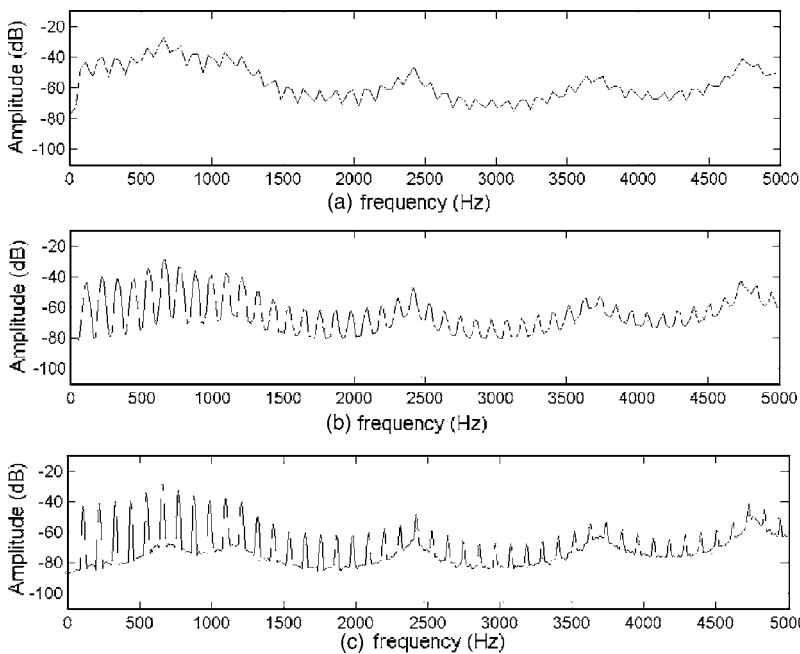
FIG. 10. Averaged spectra for 110 Hz synthesized vowel /a/ with 2% glottal noise and window lengths of (a) 256, (b) 512, and (c) 1024 samples (sampling frequency=10 kHz).

Peter J. Murphy: Rahmonic analysis of voice
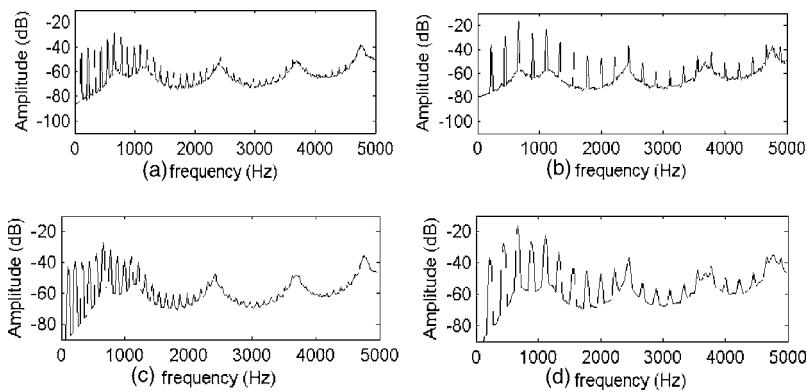
FIG. 11. Averaged, period asynchronous spectrum (1024 samples) of (a) 110 Hz and (b) 220 Hz synthesized /a/ vowel with 4% glottal noise. Averaged period-synchronous spectrum (six fundamental periods) of the same vowel [(c) 110 Hz and (d) 220 Hz].

psR1 estimate. Alternatively, the average signal amplitude over the window can be subtracted from the glottal signal as a preprocessing step.

In Fig. 8, gmHNR and psR1 are plotted against shimmer. gmHNR (estimated from the complete spectrum and the spectrum limited to the first 13 harmonics) decreases approximately linearly with doubling of shimmer levels for the glottal and speech data. The four estimates of gmHNR for a given noise level are approximately equal. Spectrally, shimmer results in a HNR that is equal at all frequencies (Murphy, 2000); it can be seen that harmonic/between-harmonic difference remains constant throughout the spectrum [Fig. 12(a)] and hence gmHNR taken from the complete spectrum and the harmonic limited spectrum are expected to be similar. As the HNR is equal at each harmonic/between harmonic, it is expected that gmHNR would equal the actual HNR (at least for the glottal signals, i.e., 40, 34, 28, 22, 16, and 10 dB for 1, 2%, 4%, 8%, 16%, 32% std. dev. shimmer levels, respectively). Although approximately 6 dB decreases are observed for the doubling of shimmer levels, gmHNR values are overestimated by approximately 8 dB. The four estimates of psR1 are also approximately equal at a given noise level. However, psR1 estimated from the harmonic limited spectrum for the glottal signal is again approximately 1 dB greater than for the harmonic limited speech spectrum. For a conventional HNR estimator, noise and shimmer levels of the same level produce the same HNR values. This contrasts HNR with gmHNR and psR1; the latter indices are different for shimmer and random noise of the same level (Fig. 3, Fig. 7, Fig. 8).

The averaged spectrum for random jitter [Fig. 12(b)] shows an early harmonic structure that deteriorates with increasing frequency. psR1 reflects these alterations in harmonic structure (Fig. 9). However, the result of averaging jittered signal spectra is quite different from the result of averaging spectra due to periodic signals plus noise. In the latter case, the harmonics approach the true signal harmonics and the noise approaches the true noise variance. In averaging the jittered signal spectra, no greater insight into quantifying the level of aperiodicity is apparent. The averaging appears to smooth out the jittered harmonics at higher frequencies and a subharmonic regime is not as evident. gmHNR and psR1 decrease nonlinearly with increased jitter for the complete spectrum and the spectrum limited to the first 13 harmonics for the glottal and speech data. The harmonic limited spectra have higher values of gmHNR and psR1 due to the relative richness of harmonic structure over this interval. Again there is an approximate 1 dB difference in psR1 (glottal—harmonic limited) compared to psR1 (speech—harmonic limited), which is attributed to the dc term of the glottal signal. Despite the strong correlations between gmHNR and psR1 for jittered signals, caution should be taken when interpreting R1 and gmHNR relationships in this case. For nonaveraged spectra of jittered signals, a subharmonic regime is usually in evidence (cf. Murphy, 2000), which may contribute differently to gmHNR and psR1 and may also increase the amplitude of the second rahmonic. Broadening of the bandwidths of the low frequency harmonics may also shift the first rahmonic location slightly. A formal study of these issues is required.

B. First rahmonic amplitude and human voice signals

A number of studies cited in the introduction (Heman-Ackah *et al.*, 2002; Hillenbrand *et al.*, 1994; Dejonckere and Wieneke, 1994; Awan and Roy, 2006) state that the first cepstral peak (R1) provides a superior acoustic index (when compared with indices such as jitter, shimmer, and various spectral tilt and noise measures) for differentiating between normal and, pathological or breathy, voice quality. It has been shown in the present development that R1 provides an indication of the geometric mean harmonics-to-noise ratio, i.e., it provides an overall index of the richness of the harmonic spectrum in dB. The synthesis shows that noise and shimmer increase between-harmonic amplitudes in the averaged spectra and hence lead to reduced values of R1. Shim-
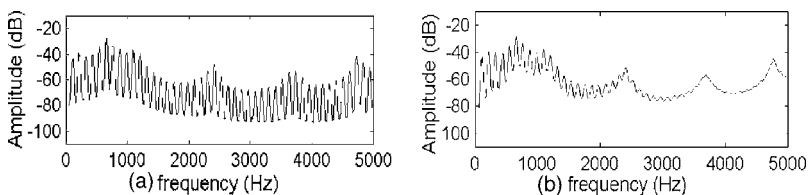
FIG. 12. Averaged, period asynchronous spectrum (512 samples) of synthesized /a/ vowel with (a) 2% shimmer and (b) 1% jitter.

mer is shown to influence R1 in a different manner from noise, as expected for a gmHNR, as detailed in the previous section. Jitter deteriorates the existing harmonic structure and increases the between-harmonics, leading to a reduced value of R1. Hence aperiodicity in the temporal waveform is reflected in the spectrum leading to a reduction in R1. Although this is also the case for a HNR estimator, the exact nature of the changes in the indices is different due to the geometric mean inherent in the R1 measure.

Because R1 provides average information regarding harmonics-to-noise ratio components in the dB spectrum, it perhaps provides information more closely associated with spectrographic (logarithmic display of amplitude) evaluation of voiced speech containing glottal noise than a conventional HNR estimator. The spectrogram displays the speech signal in terms of a gray scale representing logarithm intensity, hence when an observer makes an estimate regarding the HNR of the signal, he or she is forming a judgment based on logarithmic data. The fact that the index provides an average of logarithmic values effectively gives greater weight to the higher frequency noise components in the spectrum, which may also make the measure more perceptually relevant than a conventional HNR estimator.

A further aspect of the spectrum, not investigated through the present synthesis, concerns spectral slope. A primary correlate of breathy voice quality is spectral slope, typically given as -18 dB/octave compared to -12 dB/octave for normal voice (Childers and Lee, 1991). This slope will not affect the calculation of a conventional HNR (as long as the harmonic level exceeds the noise level in the averaged spectrum). However, based on the characterization of R1 being directly proportional to an average of dB HNRs at harmonic/between-harmonic locations spectral slope due too the glottal source will influence R1, even if the underlying signals have equal HNR values. This may be observed by directly substituting appropriate values into Eq. (9) or equivalently it is noted that a breathy voice is characterized by a more sinusoidal flow waveform with corresponding increased relative first harmonic and reduced higher harmonics. Due to the reduced periodicity in the dB spectrum, a reduced R1 value results independent of noise and perturbation considerations. This inference suggests another possible reason as to why R1 provides a better indicator of voice quality, when compared against other acoustic indices, in the studies cited.

An additional characteristic of R1 that may account for the strong performance of the measure as an indicator of voice quality (Hillenbrand *et al.*, 1994; Dejonckere and Wieneke, 1994; Heman-Ackah *et al.*, 2002 and Awan and Roy, 2006) is the relative $f0$ independence of the measure when compared to a conventional HNR estimate in a period-asynchronous analysis. Hillenbrand *et al.* (1994) and Heman-Ackah *et al.* (2002) use 25.6 ms sampled at 10 kHz, Dejonckere and Wieneke (1994) use 100 ms sampled at 20 kHz and Awan and Roy (2006) use 41 ms sampled at 25 kHz, i.e., the analyses are period asynchronous and, hence, in the case of R1 in the present synthesis, less sensitive to variation with $f0$.

VI. CONCLUSION

A theoretical analysis of the cepstrum of voiced speech containing aspiration noise predicts that the first rahmonic amplitude, R1 is directly proportional to the geometric mean harmonics-to-noise ratio (gmHNR, where gmHNR is defined as the mean of the individual harmonic to between harmonic ratios in dB). Tests on the synthesis of glottal and speech signals containing glottal noise validate the hypothesis. In addition, gmHNR (estimated directly from the speech spectrum) is shown to be approximately equal to gmHNR (estimated directly from the glottal spectrum). This approximation is also shown for R1 (speech and glottal). gmHNR and R1 are shown to decrease in a nonlinear manner with the doubling of noise levels. This is shown to be due to the fact that noise levels exceed harmonic levels (at higher frequencies) as the noise increases. Limiting the harmonics in the spectrum causes the response to the doubling of noise levels to become linear (i.e., higher harmonics are no longer exceeded in amplitude by the noise level). Limiting the spectrum to a set number of harmonics all alleviates the $f0$ dependence of the measure in a pitch synchronous analysis. The responses of psR1 and gmHNR to shimmer and jitter, respectively, are also shown to be highly correlated.

A number of reasons are inferred as to why this single acoustic index provides a strong gross correlate of voice quality: (a) the relative $f0$ independence characteristic when extracted via a period-asynchronous analysis, (b) the relative vocal tract independence of the measure, (c) greater weight is given to higher frequencies (when compared with a conventional HNR), and (d) the glottal source spectral slope sensitivity property. This is not to suggest, however, that a true multidimensional acoustic characterization of voice quality is not required. In contrast to a single global indicator, finding accurate measures that reflect specific aspects of voice quality remains an important research goal. The harmonic limited period synchronous R1 measure extracted from the dB spectrum of voiced speech shows promise for quantifying glottal signal-to-noise attributes.

[1] For completeness it is mentioned that an alternative signal-to-noise estimation technique using a region centered on the first rahmonic peak is given in Yegnanarayana *et al.* (1998). However, the cepstral processing is only used as a preprocessing step to isolate the harmonics for their recursive algorithm used in HNR estimation.

[2] In reality, signal energy is estimated at harmonic locations and noise energy is estimated at "between harmonic" locations and hence Eq. (10) is only approximately equal to Eq. (9).

[3] In de Krom (1993) and Qi and Hillman (1997) it appears that the obtained signal to noise index is also extracted directly from the power spectrum in dBs, i.e., an average of dB values (of each harmonics-to-noise ratio estimate at a given frequency) is estimated and hence is equivalent to gmHNR [Eq. (9)]. Therefore comments in Secs. V and VI regarding gmHNR are also attributable to the indices derived from these techniques. These methods can also be used to extract a conventional HNR estimate, which sums

the signal and noise energies in linear scale before taking the logarithm of their quotient.

[4]This is also the case for a conventional HNR measure but the overestimates of the higher harmonics have relatively little effect because the ratio is dominated by the low frequency energy.

Alku, P. and Vilkman, E. (**1996**). "Comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers," Folia Phoniatr Logop **48**, 240–254.

Awan, S. N. and Roy, N. (**2006**). "Toward the development of an objective index of dysphonia severity: A four factor acoustic model," Clin. Linguist. Phonetics **20**, 35–49.

Childers, D. G. and Lee, C. K. (**1991**). "Vocal quality factors: analysis, synthesis and perception," J. Acoust. Soc. Am. **90**, 2394–2410.

de Krom, G. (**1993**). "A cepstrum based technique for determining a harmonics-to-noise ratio in speech signals." J. Speech Hear. Res. **36**, 254–266.

Dejonckere, P. and Wieneke, G. H. (**1994**). "Spectral, cepstral and aperiodicity characteristics of pathological voice before and after phonosurgical treatment," Clin. Linguist. Phonetics **8**, 161–169.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers; Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Herman-Ackah, Y., Michael, D. D., and Goding, Jr., G. S. (**2002**). "The relationship between cepstral peak prominence and selected parameters of dysphonia," J. Voice **16**, 20–27.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech Hear. Res. **37**, 769–777.

Kasuya, H., Ogawa, S., Mashima, K., and Ebihara, S. (**1986**). "Normalized noise energy as an acoustic measure to evaluate pathologic voice," J. Acoust. Soc. Am. **80**, 1329–1334.

Klatt, D. and Klatt, L. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Koike, Y. (**1986**). "Cepstrum analysis of pathological voices," J. Phonetics **14**, 501–507.

Koike, Y. and Kohda, J. (**1991**). "The effect of vocal fold surgery on the speech cepstrum," in *Vocal Fold Physiology: Acoustic, Perceptual and Physiologic Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego).

Ladefoged, P. and Antonanzas-Barroso, N. (**1985**). "Computer measures of breathy voice quality," UCLA Working Papers in Phonetics **61**, 79–86.

Murphy, P. J. (**1999**). "Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis," J. Acoust. Soc. Am. **105**, 2866–2881.

Murphy, P. J. (**2000**). "Spectral characterisation of jitter, shimmer and additive noise in synthetically generated voice signals," J. Acoust. Soc. Am. **107**, 978–988.

Noll, AM. (**1967**). "Cepstrum pitch determination," J. Acoust. Soc. Am. **41**, 293–309.

Oppenheim, AV., Schafer, R. W., and Stockham, (Jr.), TG. (**1968**). "Nonlinear filtering of multiplied and convolved signals," Proc. IEEE **56**, 1264–1291.

Qi, Y. and Hillman, RE. (**1997**). "Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals," J. Acoust. Soc. Am. **102**, 537–543.

Rabiner, L. and Schafer, R. (**1978**). *Digital Processing of Speech Signals* (Prentice–Hall, Englewood Cliffs, NJ).

Rosenberg, A. (**1971**). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**, 583–590.

Yegnanarayana, B., d'Alessandro, C., and Darsinos, V. (**1998**). "Decomposition of the speech signal into deterministic and stochastic components," IEEE Trans. Speech Audio Process., **6**, 1–11.

Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels

Deniz Başkent[a]

House Ear Institute, Department of Auditory Implants, 2100 West Third Street,
Los Angeles, California 90057

Speech recognition by normal-hearing listeners improves as a function of the number of spectral channels when tested with a noiseband vocoder simulating cochlear implant signal processing. Speech recognition by the best cochlear implant users, however, saturates around eight channels and does not improve when more electrodes are activated, presumably due to reduced frequency selectivity caused by channel interactions. Listeners with sensorineural hearing loss may also have reduced frequency selectivity due to cochlear damage and the resulting reduction in the nonlinear cochlear mechanisms. The present study investigates whether such a limitation in spectral information transmission would be observed with hearing-impaired listeners, similar to implant users. To test the hypothesis, hearing-impaired subjects were selected from a population of patients with moderate hearing loss of cochlear origin, where the frequency selectivity would be expected to be poorer compared to normal hearing. Hearing-impaired subjects were tested for vowel and consonant recognition in steady-state background noise of varying levels using a noiseband vocoder and as a function of the number of spectral channels. For comparison, normal-hearing subjects were tested with the same stimuli at different presentation levels. In quiet and low background noise, performance by normal-hearing and hearing-impaired subjects was similar. In higher background noise, performance by hearing-impaired subjects saturated around eight channels, while performance by normal-hearing subjects continued to increase up to 12–16 channels with vowels, and 10–12 channels with consonants. A similar trend was observed for most of the presentation levels at which the normal-hearing subjects were tested. Therefore, it is unlikely that the effects observed with hearing-impaired subjects were due to insufficient audibility or high presentation levels. Consequently, the results with hearing-impaired subjects were similar to previous results obtained with implant users, but only for background noise conditions. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354017]

I. INTRODUCTION

Two factors that affect speech recognition by listeners with sensorineural hearing loss (SNHL) considerably are reduced audibility, due to elevated hearing thresholds, and suprathreshold deficits. In his model for recognition of speech in noise by hearing-impaired (HI) listeners, Plomp (1978) termed these factors as "attenuation" and "distortion," respectively. The attenuation problem can mainly be solved by amplification of sound to levels above the auditory thresholds. The solutions to the distortion problem are, however, not as straightforward. One form of distortion is related to reduced frequency selectivity, commonly seen in moderate to profound hearing loss of cochlear origin. The main cause for reduced frequency selectivity is thought to be damage in outer hair cells (OHCs). OHCs are believed to be an essential part of the active mechanism in cochlea that gives rise to cochlear nonlinearities such as the basilar membrane compression (Oxenham and Plack, 1997). The cochlear nonlinearities contribute to sharp tuning of the auditory filters

(AFs) in normal hearing (NH). When there is damage in the cochlea, particularly in the OHCs, the nonlinearities might be reduced and the AFs might be broadened (Glasberg and Moore, 1986; Baker and Rosen, 2002).

Festen and Plomp (1983) observed that speech recognition by HI listeners in quiet was mainly affected by audiometric thresholds, while speech recognition in noise was affected by frequency resolution. The correlation between reduced frequency resolution and poor speech intelligibility in noise was later confirmed by other studies (Stelmachowicz *et al.*, 1985; Noordhoek *et al.*, 2000). In a study by Leek and Summers (1996), listeners with broad AFs were observed to need higher spectral contrast for discrimination of vowel-like stimuli presented in noise. It was hypothesized that a reduced SNR in the internal auditory representation of speech might be an explanation for the detrimental effects of reduced frequency resolution. Some studies also showed a correlation in quiet listening conditions; Turner and Henn (1989) observed that input filter patterns, a measure of frequency resolution, were correlated with recognition of vowels, and Henry *et al.* (2005) observed that the capability of resolving spectral peaks of rippled noise was correlated with recognition of vowels and consonants.

---

[a] Present address: Starkey Hearing Research Center, 2150 Shattuck Avenue, Suite 408, Berkeley, California 94704. Electronic mail: deniz_baskent@starkey.com

Reduced frequency selectivity of hearing impairment was simulated with NH listeners as well. When speech was degraded by spectral smearing to simulate broadened AFs, speech recognition by NH listeners dropped, and the presence of the background noise amplified the detrimental effects (ter Keurs *et al.*, 1992, 1993; Baer and Moore, 1993; Boothroyd *et al.*, 1996).

Another method widely used to systematically explore the effects of temporal and spectral degradations on perception of speech is the noiseband vocoder (Shannon *et al.*, 1995; Xu *et al.*, 2005). In vocoder processing, speech signal is filtered into a number of spectral bands and narrow bands of noise are modulated with the envelopes extracted from individual bands. The final product is a synthesis of modulated noise bands. The processed speech has only crude spectral and temporal elements of the input speech. Cochlear implants (CIs) similarly deliver mainly speech envelope information, by modulating the current that stimulates the auditory nerve, and the gross spectral information, by stimulating distinct tonotopic places along the cochlea. Due to the similarity in the processing, despite the unknown percept of electrical stimulation, the noiseband vocoder has also been used to simulate CI processing with NH listeners.

Using a noiseband vocoder, spectral resolution of speech stimulus can be manipulated by changing the number of spectral channels. When vocoder speech was presented to NH subjects, performance increased as a function of the number of spectral channels. When CI users were tested under similar conditions, however, their performance usually saturated by 4–8 electrodes, and did not improve as more electrodes were activated (Fishman *et al.*, 1997; Friesen *et al.*, 2001; Garnham *et al.*, 2002). This finding suggests that, similar to HI listeners, implant users also have reduced frequency selectivity; they were unable to make use of the fine spectral resolution provided by the larger number of electrodes. However, unlike HI listeners, the main limiting factor for frequency resolution by implant users is believed to be channel interactions, caused by the summation of electric current fields or stimulation of the same nerve population by electrical pulses sent from different electrodes (Shannon, 1983).

Turner *et al.* (1995, 1999) used vocoder processing to explore the effects of reduced temporal or spectral resolution due to hearing impairment on perception of consonants in quiet. In the earlier study (Turner *et al.*, 1995), when a single-channel noiseband vocoder was used to eliminate all spectral cues, HI and NH listeners performed similarly. It was concluded that HI listeners were able to receive temporal information similar to NH listeners. In the second study, Turner *et al.* (1999) varied the number of channels from one to eight. The hypothesis was if the reduced frequency resolution of HI listeners was a limiting factor, the HI performance would reach an asymptote at a smaller number of channels compared to NH listeners, as it was the case with CI users. Performance by both subject groups continued to increase as a function of the number of spectral channels. However, except for the one-channel condition, performance by HI listeners was lower than NH listeners. Turner *et al.* concluded that even though HI listeners were able to make use of the temporal cues with one-channel processing, they were unable to combine the information from multiple channels as efficiently as NH listeners.

In the present study, it is hypothesized that the limitation can be observed if the experimental conditions of Turner *et al.* (1999) were expanded. There are several factors that might have prevented Turner *et al.* (1999) from fully observing the limiting effects of reduced spectral resolution on the speech recognition performance of HI listeners. The present study was designed by improving several factors: (1) In the Turner *et al.* (1999) study, the number of channels was increased only up to eight channels. Friesen *et al.* (2001) observed most difference in the performance by NH and CI subjects for number of channels higher than seven. For smaller number of channels, performance by implant users was lower than the performance by NH listeners, except for the single-channel condition, similar to findings by Turner *et al.* (1999). In the present study, the number of channels was varied from 2 to 40. (2) Previous studies showed strongest effects of reduced frequency selectivity due to hearing impairment on the perception of speech in noise (Festen and Plomp, 1983; Horst, 1987; Leek and Summers, 1996). Therefore, background noise of varying levels was added to the experimental conditions. (3) Perception of vowels has generally been observed to be more sensitive to spectral manipulations than consonants (Boothroyd *et al.*, 1996; ter Keurs *et al.* 1992; Turner and Henn, 1989). Therefore, the effects of changing the number of spectral channels could be stronger with vowel perception. In the present study, vowels were used as stimuli as well as consonants. (4) As mentioned above, audibility is one of the main factors affecting speech perception performance by HI listeners (i.e., Plomp, 1978). If it is not carefully controlled for, it might be difficult to separate the effects of suprathreshold deficits from the effects of audibility. Turner *et al.* (1999) maximized the audibility for HI listeners and used high presentation levels with NH listeners. In the present study, the audibility was similarly maximized for HI listeners. The presentation levels for NH listeners, however, were set at different levels to identify possible audibility effects. (5) It is also important to select the appropriate level of hearing loss. In the study by Turner *et al.* (1999), the baseline consonant recognition scores with unprocessed stimuli were considerably low with two out of six HI subjects. One of the two subjects with low scores had a severe hearing loss at high frequencies. Therefore, this subject probably did not have access to spectral information higher than 2 kHz, a region important for perception of consonants. In the present study, the inclusion criteria comprised having relatively flat hearing loss, from 50 to 60 dB HL, for frequencies up to 6 kHz. OHC damage was further confirmed by otoacoustic emission (OAE) measurements. To reduce possible audibility effects at higher frequencies, vocoder processing was limited to frequencies lower than 6 kHz.

The second purpose of the present study was to compare the results with HI listeners to the results with CI users (Friesen *et al.*, 2001).

HI listeners have inherently reduced frequency selectivity in the auditory system while the frequency selectivity of

TABLE I. Information about the HI subjects.

Subject	Age	Cause of the hearing loss	Age at diagnosis of the hearing loss (years)	Hearing aid user	Hearing thresholds (dB HL) of the test ear at the audiometric frequencies (Hz)					
					250	500	1000	2000	4000	6000
S1	34	German Measles	Shortly after birth	N	55	60	55	60	60	70
S2	62	Presbycusis	54	Y	40	50	60	50	50	65
S3	52	Streptomycin	8	Y	30	40	50	55	60	55
S4	63	Presbycusis	58	Y	45	55	55	50	55	65

CI users is thought to be mainly limited by channel interactions of the device processing. Therefore, many researchers proposed ways to minimize channel interactions in order to increase the effective number of spectral channels. For example, the "Continuous Interleaved Sampling" strategy, which delivers current pulses interleaved in time, was developed after strong channel interactions were observed with simultaneous activation of multiple electrodes (Wilson et al., 1991). Electrode designs were improved to minimize current spread (e.g., Gstoettner et al., 2001). Also different modes of stimulation, such as bipolar (BP) or tripolar (TP), have been suggested as an alternative to monopolar (MP) mode (i.e., Bierer and Middlebrooks, 2002). Many improvements in signal processing, hardware design, or stimulation methods have been shown to result in more localized stimulation patterns in physiological or psychophysical experiments. However, a beneficial effect of reduced channel interactions on the perception of speech has not been clearly demonstrated.

Studies on channel interactions may produce different results depending on the experimental design. In animal experiments, for example, the current level is usually set at a fixed dB level above threshold for all conditions. Stimuli are not loudness-balanced as this would require extensive training, and the dynamic range cannot be measured as the animals are incapable of reporting the maximum acceptable loudness levels. In BP and TP modes, the return electrodes are closer to the active electrode. Therefore, at the same current level, MP configuration would produce a wider stimulation pattern than the BP and TP configurations (Bierer and Middlebrooks, 2002). A higher degree of channel interaction was also observed with the MP configuration (Bierer and Middlebrooks, 2004). With human subjects, on the other hand, the stimuli can be loudness-balanced and the dynamic range can be measured. Usually, a small current is sufficient to produce the same loudness in MP mode as the BP or TP modes, as a larger population of nerves is stimulated in the MP mode due to the far positioning of the return electrode. When Kwon and van den Honert (2006) normalized the data using loudness-balancing and dynamic range measurements, the channel interaction patterns were similar for MP and BP modes. This observation might provide an explanation for why different stimulation modes do not always produce different speech recognition performance (i.e., Pfingst et al., 1997; 2001).

The results with other measures of channel interactions have been mixed as well. Cazals et al. (1990) have not found a correlation between forward-masking patterns and speech

recognition performance by CI users. Chatterjee and Shannon (1998) observed that one subject with the lowest sentence recognition scores had masking patterns with greatest dependence on masker level and on probe delay. Throckmorton and Collins (1999) measured channel interactions using electrode discrimination and forward masking and measured speech intelligibility with a number of stimuli of varying complexity. The results from both methods were correlated with some, but not all, measures of speech recognition. Stickney et al. (2006) showed that high levels of electric field interactions were correlated with low speech intelligibility performance for a simultaneous speech processing strategy (Simultaneous Analog Stimulation), but not for a sequential speech processing strategy (Continuous Interleaved Sampling).

The absence of a robust correlation between channel interactions (as measured in psychophysical or physiological studies) and speech recognition may imply that there might be other factors limiting the frequency selectivity, presumably originating in the auditory system, in addition to channel interactions caused by implant processing. In the present study, subjects with moderate SNHL also served as a model for an auditory system with inherently reduced frequency selectivity. It was explored if reduced frequency resolution in the auditory system could affect the performance in a manner similar to reduced frequency resolution observed with CI users.

II. METHODS

A. Subjects

Four HI subjects, aged 34–63 years (with an average of 52.75 years), participated in the study. All HI subjects were reported to have hearing loss of cochlear origin by their clinicians. All subjects other than S1 had symmetrical hearing loss. In the nontest ear, S1 had a mild hearing loss for low- to mid-frequencies, and a moderate hearing loss for higher frequencies; audiometric thresholds were rising from 35 to 25 dB HL for frequencies less than 1 kHz, sloping down from 25 to 55 dB HL for frequencies up to 4 kHz, and flat at 55 dB HL at higher frequencies. Detailed information about the HI subjects can be found in Table I.

Each subject was tested on one ear only. Test ears met the following criteria: (1) Relatively flat audiogram: A prescription method would have allowed customized amplification; however, it would also alter the spectral shape of speech. To prevent such additional factors that might affect

the results, HI subjects with relatively flat audiogram were selected and the speech was made audible at all frequencies by changing the overall presentation level only.

(2) Pure tone thresholds around 50–60 dB HL for the audiometric frequencies of 250 Hz to 6 kHz: Moore (1996) suggested that for hearing loss up to 45 dB HL audibility is the most important factor for perception of speech; for higher degrees of hearing loss, effects of suprathreshold deficits can also be observed (Carney and Nelson, 1983). van Tasell (1993) suggested that a hearing loss less than 60 dB HL is generally associated with OHC loss. For these reasons, HI subjects were selected from a population of patients with hearing loss varying between 50–60 dB HL. These levels were optimal for the present study; the potential audibility effects could be minimized by proper amplification, and the suprathreshold deficits, presumably due to OHC loss, could still be observed.

(3) Nonfunctioning outer hair cells: One of the main assumptions in the present study is that the cochlear damage is primarily in the OHCs. OAEs are generally associated with the active mechanism in the cochlea and thought to represent the OHC function (Norton, 1992). Therefore, damage in OHCs was confirmed with (the absence of) OAEs.

(4) No dead regions: If there is additional damage in the inner hair cell or auditory nerve, the speech intelligibility performance would further be affected, and in possibly different ways depending on the pathology. To rule out dead regions, the Threshold Equalizing Noise (TEN) test (Moore *et al.*, 2000) was used. There was no elevation in the masked pure-tone thresholds with the TEN test for three subjects. The test was not applicable for subject S1 as the subject found the masking noise levels of the test uncomfortably high.

As a control group, five NH subjects, aged 27–35 years (with an average of 29.00 years), were tested with the same experimental conditions. NH subjects had thresholds better than 20 dB HL at audiometric frequencies from 250 Hz to 8 kHz. The right ear was chosen as the test ear for all NH subjects.

All subjects were native speakers of American English. The average pure tone thresholds of the test ears for both subject groups are shown in Fig. 1.

B. Stimuli

The speech recognition task was identification of medial vowels and consonants. Vowel stimuli (Hillenbrand *et al.*, 1995) consisted of twelve medial vowels presented in an /h/-vowel-/d/ context (heed, hid, head, had, hod, hawed, hood, who'd, hud, heard, hayed, hoed). The phonemes were spoken by 3 female and 3 male talkers and a total of 72 tokens were presented for each condition. Chance level on this test was 8.33% correct. Consonant stimuli (Shannon *et al.*, 1999) consisted of 20 medial consonants (/b tʃ d ð f g ʒ k l m n p r s ʃ t v w j z/), presented in an /a/-consonant-/a/ context. The phonemes were spoken by 2 female and 2 male talkers and a total of 80 tokens were presented for each condition. Chance level for this test was 5% correct.

For all stimuli, percent correct scores (PCS) were corrected for chance using the equation
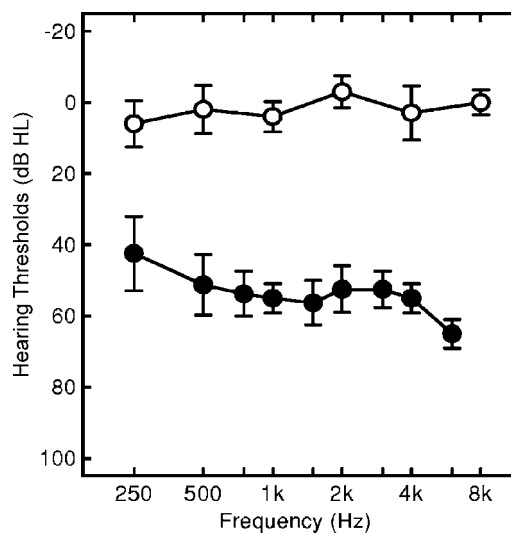
FIG. 1. Average pure tone hearing thresholds of the subjects, shown in dB HL. The open symbols show the thresholds of NH listeners for the audiometric frequencies from 250 Hz to 8 kHz, and the filled symbols show the thresholds of the HI listeners for the audiometric frequencies from 250 Hz to 6 kHz. Error bars show one standard deviation.

$$PCS_{corrected\ for\ chance} = (PCS_{raw}\text{-Chance level})/$$
$$(100\text{-Chance level}) \times 100.$$

In the corrected scores, a performance at 0% represents a performance at chance level.

In all experiments, a broadband noise with speech-shaped spectrum (SSN) was used. SSN was generated by filtering white noise using a Butterworth lowpass filter with a rollover of −12 dB\octave and a cutoff frequency of 800 Hz. The same SSN was also presented to the contralateral ear at a level 25 dB lower than the stimulus presentation level, to mask possible acoustic leakage to the nontest ear. The main motivation was to prevent one subject (S1) with asymmetrical hearing loss from listening with the better ear.

A Bruel and Kjaer artificial ear was used for calibration of the experimental setup. A reference tone signal at 1 kHz was generated with a total rms value of 65 dBA at the headphone output. The reference tone was used in setting the presentation level of stimuli to correct dBA values. The setting of the Crown Amplifier was fixed for the maximum presentation level of 95 dBA and a Tucker-Davis attenuator was used to produce the desired presentation levels. For the conditions with background noise, the speech and noise levels were proportionally adjusted to have the appropriate SNR, by using total rms values. The level of the combined signal was set to the presentation level. As a result, speech level decreased slightly as the SNR decreased.

C. Experimental procedure

The stimuli were processed with the noiseband vocoder (Shannon *et al.*, 1995) in real-time and presented using the CAST software created by Qian-Jie Fu at the House Ear Institute.

The overall frequency range of speech was limited to 200–6000 Hz. First, the stimulus was processed into a number of spectral channels using a set of Butterworth bandpass

filters with a rollover of −24 dB\octave. The cutoff frequencies were determined by logarithmically dividing the input spectral range of 200–6000 Hz. The envelopes were extracted using half-wave rectification followed by a Butterworth lowpass filter with a cutoff frequency of 160 Hz and a rollover of −24 dB\octave. These bands were the analysis bands of the vocoder. The carrier noise bands were obtained by filtering wideband noise with the same set of bandpass filters. The envelopes from the analysis bands were used to modulate the carrier noise bands. In the last stage, modulated noise bands were combined to produce the processed stimulus, which at this point had only the coarse spectral and temporal information of the original stimulus and most fine structure was absent.

Phoneme recognition by NH and HI listeners was measured with unprocessed original stimuli and as a function of the number of spectral bands (2, 4, 6, 8, 10, 12, 16, 24, and 40 channels) of the vocoder, at varying background noise levels (quiet, SNR=10 dB, SNR=0 dB for vowels and consonants, and SNR=−5 dB for vowels only). The presentation level for the HI group was chosen to provide maximum audibility and comfort. NH listeners were tested at three presentation levels (experiments 1–3). The selection of the presentation levels is explained in detail in next section.

In experiment 4, an attempt was made to simulate the broad AFs with NH subjects. Dreschler and Plomp (1980) suggested two possible types of deterioration in the AF shape in SNHL: (1) critical bandwidth is broader; (2) slopes of the tuning curves are shallower. In experiment 2, the second type of deterioration was simulated by changing the filter slopes of the vocoder bandpass filters from −24 dB\octave to −6 dB\octave, while the cutoff frequencies remained the same. Shallower filter skirts resulted in considerable overlap between adjacent bands and produced spectrally smeared stimuli.

The speech stimuli were presented monaurally over Sennheiser HDA 200 headphones in a double-wall soundproof booth. A menu with the list of all possible phonemes was shown on the screen during testing. The subject identified the phoneme that was presented by selecting the appropriate entry in the menu using the mouse. The presentation order of individual tokens in each condition and the testing order of different conditions were randomized to minimize learning effects. As an additional caution, HI subjects were given one practice session of 2 h with similar testing procedure, where feedback was also provided, prior to actual data collection. NH subjects were familiar with noiseband vocoder technique from previous studies and therefore were not given a practice session. All subjects were allowed a preview of the stimuli at the beginning of each test.

NH subjects were tested once with each condition. Most HI subjects were tested more than once depending on subject's availability. Subjects S1 and S4 were tested twice for each condition. Subject S2 was tested three times with consonants for the noise conditions of 10 dB and 0 dB SNR, and three times with vowels for the noise conditions of 0 dB and −5 dB SNR.
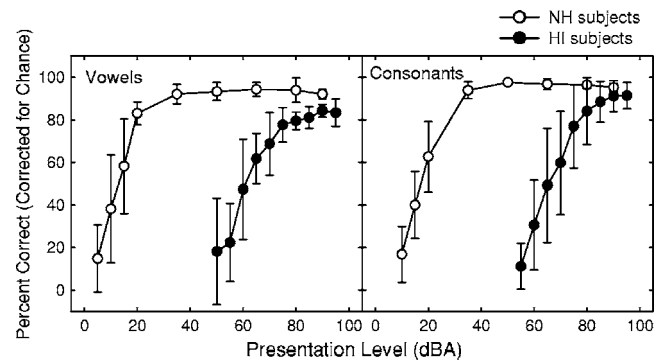
FIG. 2. Performance intensity (PI) functions. The average percent correct scores are plotted as a function of the stimulus presentation level. Open and filled symbols show the PI functions for NH and HI subjects, respectively. The stimuli were vowels and consonants, as shown in the left and right panels, respectively. The error bars show one standard deviation.

III. AUDIBILITY AND PRESENTATION LEVELS

The comfortable level for conversational speech is around 60–70 dB SPL. However, if speech was presented at an overall presentation level of 65 dB SPL in the present study, HI listeners would have insufficient audibility due to their elevated hearing thresholds. The presentation level for HI users, therefore, had to be set at higher levels to maximize audibility. On the other hand, the highest level could not exceed 100 dB SPL or so, as the stimuli would become uncomfortably loud due to loudness recruitment. For a fair comparison, the comfortable level of 65 dB SPL was not used with NH listeners either. Instead, the presentation levels were selected to be similar, either in sensation or in absolute levels, to the levels used with HI subjects.

The present study controls for the level effects using performance intensity (PI) functions, which were obtained by measuring percent correct scores with the test stimuli as a function of the overall presentation level. This method was preferred over using an amplification based on pure-tone thresholds or Speech Intelligibility Index (SII, ANSI S3.5-1997), as the PI functions provided a functional measure of how well the subjects performed with the specific stimuli used in the present study. The PI functions measured with unprocessed vowels and consonants are shown in the left and right panels, respectively, in Fig. 2. The open and filled symbols show the percent scores, corrected for chance, for NH and HI listeners, respectively.

The figure shows that NH listeners had a wide dynamic range for optimal presentation levels. From 35–40 dBA to 90 dBA, speech intelligibility was best and the loudness was reported to be comfortable. 50% of the peak PI levels (PI-50) was observed around 10–20 dBA. In contrast, speech intelligibility by HI listeners was best for high presentation levels, ranging from 80 to 95 dBA. For higher levels the stimulus was uncomfortably loud. PI-50 was observed for presentation levels around 60–65 dBA. As a result, the optimal listening levels were limited to the range from 80 to 95 dBA for HI subjects, significantly smaller than the range of optimal levels for NH listeners.

Note that even at the loudest presentation levels, the vowel recognition by HI listeners was slightly lower than the

Deniz Başkent: Frequency selectivity in sensorineural hearing loss
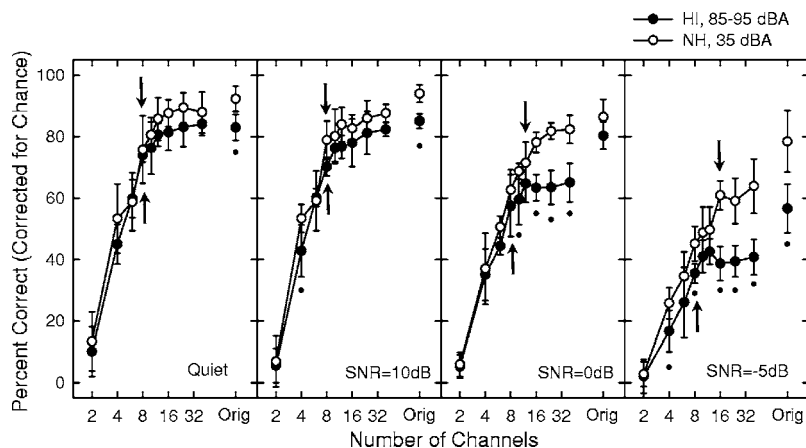
FIG. 3. Average vowel recognition scores, corrected for chance, shown as a function of the number of spectral channels. From left to right panel the background noise changes from quiet to SNR=−5 dB. The open and filled symbols show the performance by NH and HI listeners, respectively. The stars under the scores show the significant difference in the performance ($p < 0.05$) by NH and HI listeners based on a posthoc Tukey multiple comparisons test, following a two-way mixed ANOVA. The performance by NH and HI groups for the unprocessed stimuli was compared using a t-test ($p < 0.05$). The arrows above and below the data indicate the number of channels where the performance by NH and HI listeners, respectively, asymptoted.

NH listeners. When the vowel confusion matrices were analyzed, it was observed that all HI subjects and two out of five NH subjects had difficulty distinguishing the phoneme pair of "hod/hawed." Turner and Henn (1989) showed that recognition of vowels by HI listeners could be low if the formants of the vowels are similar. As the vowels used in their study were synthetically produced, the duration, which can be a robust cue for differentiating vowels with similar formants, was the same for all phonemes. For the pair of "hod/hawed," the formants are similar, and the duration cue is also not useful as both phonemes are of long duration. Differentiating these phonemes might have been a more difficult task for HI listeners than the NH listeners. The confusion caused by this one vowel pair could be an explanation for the slight difference observed in the performance by NH and HI listeners.

In the experiments, the presentation level for HI subjects was limited to the range from 85 to 95 dBA where all subjects had the best speech recognition. Subjects were given the flexibility to select the most comfortable level within this range. In general, higher levels were preferred for consonants than vowels. Subject S1, who had the poorest pure tone thresholds among all subjects, preferred louder levels.

NH subjects were tested at three presentation levels selected by comparing the PI functions of the two subject groups:

(A) Experiments 1 and 4: Comparable to HI subjects in sensation level (SL) with respect to PI-50; 35 dBA for vowels and 50 dBA for consonants.

The presentation levels of 85–95 dBA, at which the HI subjects were tested, were 25–30 dB higher than the presentation level at PI-50. For NH subjects, PI-50 was observed around 10 dBA for vowels, and around 20 dBA for consonants. The PI-50 presentation levels were defined as the reference value of 0 dB SL for each subject group. The presentation levels used for NH subjects were, then, made comparable to levels used with HI subjects in SL, by adding 25 dB to 10 dBA for vowels, and by adding 30 dB to 20 dBA for consonants.

(B) Experiment 2: Low level with reduced audibility; 20 dBA for vowels and 30 dBA for consonants.

To see the effects of decreased audibility the presentation levels were set to only 10 dB higher than the levels at PI-50. As a result, the levels were lower by 15–20 dB com-

pared to experiment 1. It can be seen on the PI functions that the performance by NH subjects at these levels is poorer than the best performance (Fig. 2).

(C) Experiment 3: Comparable in absolute level, 85 dBA.

The shapes of AFs are level-dependent. As the level increases the filter shape becomes broader and more asymmetrical with an elevation on the low-frequency side (Glasberg and Moore, 2000). Therefore, even in NH subjects, frequency selectivity might be different at higher stimulus levels. A clear connection between the changes in AF shapes and overall speech perception has not been shown, yet several studies suggested that speech recognition by NH listeners might differ at high presentation levels. Studebaker *et al.* (1999), for example, observed a decrease in the intelligibility of speech in noise as the level increased from 64 to 99 dB SPL. Hornsby *et al.* (2005) observed a decrease in the transmission of consonant features at high levels. To account for potential effects of high presentation levels, NH listeners were tested at an absolute presentation level similar to the levels used with HI subjects.

IV. RESULTS

A. Experiment 1: Effect of the number of spectral channels

In experiment 1, presentation levels were 85–95 dBA for HI listeners. For NH listeners, the levels were set to 35 and 50 dBA, comparable to the levels selected for HI listeners in SL (as defined with respect to PI-50 levels), for vowel and consonant recognition tasks, respectively.

Percent correct scores for vowel recognition, averaged across subjects and corrected for chance, are plotted in Fig. 3, as a function of the number of spectral channels and for varying background noise levels. Performance by NH and HI groups is shown with open and filled symbols, respectively.

Performance by NH and HI listeners was compared using a two-may mixed Analysis of Variance (ANOVA), with the main factors of subject group and number of channels, and the interaction between the two factors. Performance by both groups increased as the number of the channels increased for all background noise conditions ($p < 0.001$). The effect of the number of channels was similarly significant for all experiments reported in the present study, and therefore

TABLE II. F values of the two-way mixed ANOVA, used for comparing the performance by NH and HI subjects in experiments 1–3. The main effect of number of channels, which was significant for all experiments ($p < 0.001$), was not included. The main effect of subject group and the interaction between group and number of channels are shown for recognition of vowels and consonants, presented in varying background noise levels.

	Vowel recognition			Consonant recognition		
	Noise	Group factor $F(1,7)$	Group-channel interaction $F(8,56)$	Noise	Group factor $F(1,7)$	Group-channel interaction $F(8,56)$
Expt 1: Comfortable level	Quiet	3.10	0.37	Quiet	24.88[b]	0.83
	SNR=10 dB	3.94	1.03	SNR=10 dB	18.30[b]	7.73[b]
	SNR=0 dB	16.11[b]	3.15[b]	SNR=0 dB	17.40[b]	8.02[b]
	SNR=−5 dB	22.33[b]	4.13[b]			
Expt 2: Low level for NH listeners	SNR=0 dB	0.02	0.75	SNR=10 dB	0.74	2.22[a]
	SNR=−5 dB	2.23	1.87	SNR=0 dB	8.75[a]	3.59[b]
Expt 3: High level for NH listeners	SNR=0 dB	1.16	1.31	SNR=10 dB	41.68[b]	5.00[b]
	SNR=−5 dB	6.73[a]	2.66[a]	SNR=0 dB	55.30[b]	3.13[b]

[a]$p < 0.05$.
[b]$p < 0.01$.

will not be mentioned in reporting the results of the following experiments for simplicity. The F values are shown in Table II for the main factor of subject group and the interaction between group and number of channels. The F values with 'a' and 'b' denote significance at the levels of $p < 0.05$ and $p < 0.01$, respectively.

Following the ANOVA, a posthoc Tukey multiple comparisons test was used to compare all scores in pairs. The Tukey comparisons for the main factor of subject group within the same number of channels were used to identify the specific conditions where the performance by NH and HI listeners differed significantly. The scores with the unprocessed stimuli were not included in the ANOVA, therefore they were compared using a t-test. In Fig. 3, the conditions where the difference in the scores was statistically significant ($p < 0.05$) are shown by the dots under the scores. The Tukey comparisons for the main factor of number of channels within NH or HI subjects were used to identify the number of channels where the performance by that subject group reached the asymptote. For each subject group, the highest percent correct scores were used in the comparisons. The lowest number of channels where the performance did not differ significantly compared to the highest percent correct scores was accepted as the saturation point. In Fig. 3, the number of channels where the performance reached the asymptote is shown by arrows above the data, for NH listeners, and below the data, for HI listeners. Table III summarizes the number of channels at the saturation point for all experiments of the present study.

There was no significant main effect of group on vowel recognition in quiet and at the low noise level of SNR =10 dB. For both NH and HI subjects, performance reached the asymptote around 8 channels (Fig. 3, left panels; Table III). However, performance by NH and HI subjects differed significantly when the background noise level was higher;

there was a significant effect of subject group ($p < 0.01$) and a significant interaction ($p < 0.01$) between group and number of channels, as shown in Table II for SNR=0 dB and SNR=−5 dB. As the noise level increased, NH listeners were able to employ a higher number of spectral channels. NH performance increased up to 12 and 16 channels for SNR=0 dB and SNR=−5 dB, respectively, while performance by HI listeners saturated at 8 channels for all noise conditions (Fig. 3, right panels; Table III).

Figure 4 shows the average percent correct scores for consonant recognition, corrected for chance, as a function of the number of spectral channels and for varying background noise levels.

There was a significant main effect of subject group on

TABLE III. The summary for the number of channels where vowel and consonant recognition by NH and HI subjects saturated.

	Vowels				Consonants		
	Quiet	SNR= 10 dB	SNR= 0 dB	SNR= −5 dB	Quiet	SNR= 10 dB	SNR= 0 dB
HI Expt 1: Comfortable level	8	8	8	8	8	8	8
NH Expt 1: Comfortable level	8	8	12	16	8	10	12
NH Expt 2: Low level	–	–	12	12	–	10	16
NH Expt 3: High level	–	–	12	12	–	12	8
NH Expt 4: Wide carrier filters	–	–	16	–	–	16	–
NH Expt 4: Wide carrier and analysis filters	–	–	24	–	–	24	–

Deniz Başkent: Frequency selectivity in sensorineural hearing loss
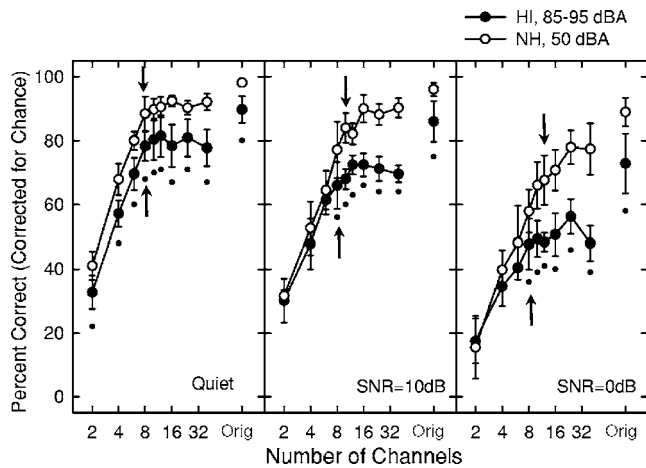
FIG. 4. Similar to Fig. 3, except for the stimuli were consonants and the background noise increases from quiet to SNR=0 dB, from left to right panels. The open and filled symbols show the scores by NH and HI subjects, respectively. The dots under the scores denote the specific conditions where the performance was significantly different ($p < 0.05$) between NH and HI listeners. The arrows show the number of channels where the performance by a specific subject group reached the asymptote.

consonant recognition in all background noise conditions ($p < 0.01$, as shown in Table II). In quiet, however, the patterns were similar as the interaction between group and channels was not significant, and performance by both groups reached the asymptote around 8 channels (Fig. 4, left panel; Table III). At higher noise levels, on the other hand, the patterns were significantly different; there was a significant interaction of group and channels at SNR=10 dB and SNR =0 dB ($p < 0.01$, as shown in Table II). As the background noise increased from SNR=10 dB to SNR=0 dB, performance by NH subjects increased up to 10 and 12 channels, respectively, while performance by HI listeners saturated by 8 channels for both noise conditions (Fig. 4, right panels; Table III).

Note that, for both vowels and consonants, there was a difference in the performance by NH and HI listeners even in the quiet and unprocessed conditions. The difference in the performance with unprocessed stimuli generally increased as the noise level increased; the drop in the performance by HI listeners was sharper than the drop in the performance by NH listeners as a function of the background noise level. When the noiseband vocoder processing was added, there was generally a further drop from the performance with the unprocessed condition, even at the highest number of channels used in the present experiment. The difference in the performance

between the 40-channel and unprocessed conditions are shown in percent correct scores in Table IV, for both subject groups and for both vowels and consonants. A paired t-test was used to determine statistical significance. The drop in the performance from the unprocessed condition to the 40-channel processing condition also increased with increasing noise, and this effect was generally more pronounced with HI listeners (Table IV).

B. Experiment 2: Effects of audibility

In experiment 2, the effect of reduced audibility on recognition of vocoder speech was explored. The presentation level for NH listeners was reduced by 15–20 dB compared to experiment 1; the vowels were presented at 20 dBA and the consonants were presented at 30 dBA. Only the noise conditions from experiment 1 where the patterns of the performance by two subject groups were significantly different were repeated. The average percent correct scores at the low presentation levels are presented in Fig. 5, superimposed with the scores from experiment 1. Figure 5(A) shows the results for vowel recognition with background noise of SNR=0 dB and SNR=−5 dB in the left and right panels, respectively. Figure 5(B) shows the scores for consonant recognition with SNR=10 dB and SNR=0 dB in the left and right panels, respectively. In each panel, open and filled circles replicate the scores by NH and HI subjects, respectively, from experiment 1. Open triangles show the scores by NH subjects when the stimuli were presented at lower levels.

The performance by NH listeners at the low presentation level and the performance by HI listeners from experiment 1 were compared using a two-way mixed ANOVA, with the main factors of subject group and number of channels, and the interaction between the two factors. The F values are shown in Table II for the main factor of group and the interaction between group and number of channels. A Tukey test was used to compare the scores with the vocoder processed conditions, and a t-test was used to compare the scores with the unprocessed conditions. The dots under the data show the conditions where the performance by NH listeners at the low presentation level and the performance by HI subjects differed significantly ($p < 0.05$). The Tukey test was also used to determine the number of channels where the performance reached the asymptote, as it was shown by arrows above the data, for NH listeners, and below the data, for HI listeners (Fig. 5).

The effect of presentation level on performance by NH listeners was explored by comparing the scores measured at

TABLE IV. The difference in performance between unprocessed and 40-channel processing conditions shown in percent correct score for vowels and consonants and for each subject group.

	Vowels				Consonants		
	Quiet	SNR= 10 dB	SNR= 0 dB	SNR= −5 dB	Quiet	SNR= 10 dB	SNR= 0 dB
NH subjects	4.40%	6.36%[a]	3.94%	14.54%[a]	6.05%[a]	5.79%[a]	11.58%
HI subjects	1.08%	2.65%[b]	15.26%[b]	15.82%	11.99%[b]	16.34%[b]	24.89%[b]

[a]Paired t-test: $p < 0.05$.
[b]Paired t-test: $p < 0.01$.
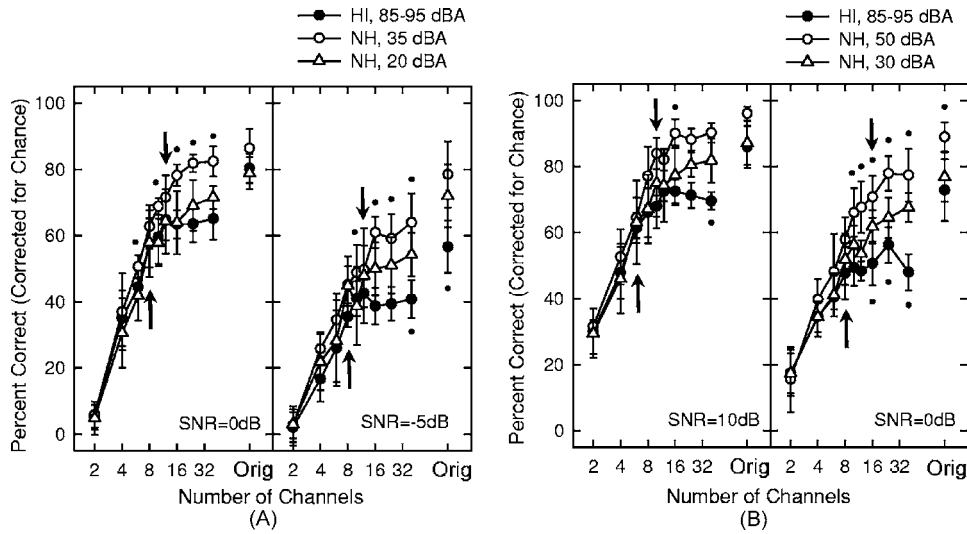
FIG. 5. Average phoneme recognition scores by NH listeners from experiment 2, where the presentation level was low, presented with the results from experiment 1, where the presentation level was comfortable. (A) shows the vowel recognition scores by HI listeners (filled circles, replicated from Fig. 3) and by NH listeners with stimuli presented at 35 dBA (open circles, replicated from Fig. 3) and at the low level of 20 dBA (open triangles). Left and right panels show the results with background noise at SNR=0 dB and SNR=−5 dB, respectively. (B) shows the consonant recognition scores by HI listeners (filled circles, replicated from Fig. 4) and by NH listeners when the stimuli were presented at 50 dBA (open circles, replicated from Fig. 4) and at the low level of 30 dBA (open triangles). Left and right panels show the results with background noise at SNR=10 dB and SNR=0 dB, respectively, The small dots above the open circles indicate the conditions where the performance by NH subjects dropped significantly as a result of reduced presentation level, shown with a Tukey test ($p < 0.05$). The small dots under the data indicate the conditions where the performance by NH subjects at the reduced presentation level was significantly different than the performance by HI subjects, shown with a Tukey test ($p < 0.05$). The conditions with unprocessed original stimuli were compared with a t-test ($p < 0.05$). The arrows above and below the data indicate the number of channels where the performance by NH and HI listeners, respectively, asymptoted.

the low presentation level of experiment 2 and the scores measured at the comfortable level of experiment 1. A two-way repeated-measures (RM) ANOVA with the main factors of presentation level and number of channels was used. The corresponding F values are presented for the main factor of presentation level and the interaction between level and number of channels in Table V. The dots above the data show the conditions where the performance by NH listeners at two presentation levels differed significantly, as shown by the Tukey or t-test ($p < 0.05$).

The results, presented in Fig. 5, show that reduced audibility generally produced lower scores than the scores at comfortable level. In unprocessed conditions, performance

by NH listeners at low presentation level was more similar to performance by HI listeners. Even with the highest number of channels, performance with processed stimuli was lower than the unprocessed conditions for both subjects groups; however, the drop in performance with vocoder processing was generally larger for HI listeners.

In the conditions with processed vowels, there was a significant main effect of presentation level on NH performance at SNR=0 dB noise level ($p < 0.05$, Table V). No significant effect of level at SNR=−5 dB, and no significant interaction at both background noise levels were observed. The lower scores by NH listeners, obtained in experiment 2 as a result of the low presentation level of 20 dBA, were not

TABLE V. F values of the two-way RM ANOVA. Performance by NH listeners at low and high presentation levels (experiments 2 and 3, respectively) was compared to performance by NH listeners at comfortable levels (experiment 1). The main effect of the presentation level and the interaction between level and number of channels are shown for recognition of vowels and consonants, presented in varying background noise levels.

	Vowel recognition			Consonant recognition		
	Noise	Level factor $F(1,4)$	Level-channel interaction $F(8,32)$	Noise	Level factor $F(1,4)$	Level-channel interaction $F(8,32)$
Expt 2: Low level for NH listeners	SNR=0 dB	15.53[a]	1.33	SNR=10 dB	2.97	1.50
	SNR=−5 dB	7.30	1.63	SNR=0 dB	10.83[a]	2.59[a]
Exp 3: High level for NH listeners	SNR=0 dB	4.12	1.23	SNR=10 dB	0.66	2.39[a]
	SNR=−5 dB	1.67	1.65	SNR=0 dB	2.77	2.41[a]

[a]$p < 0.05$.
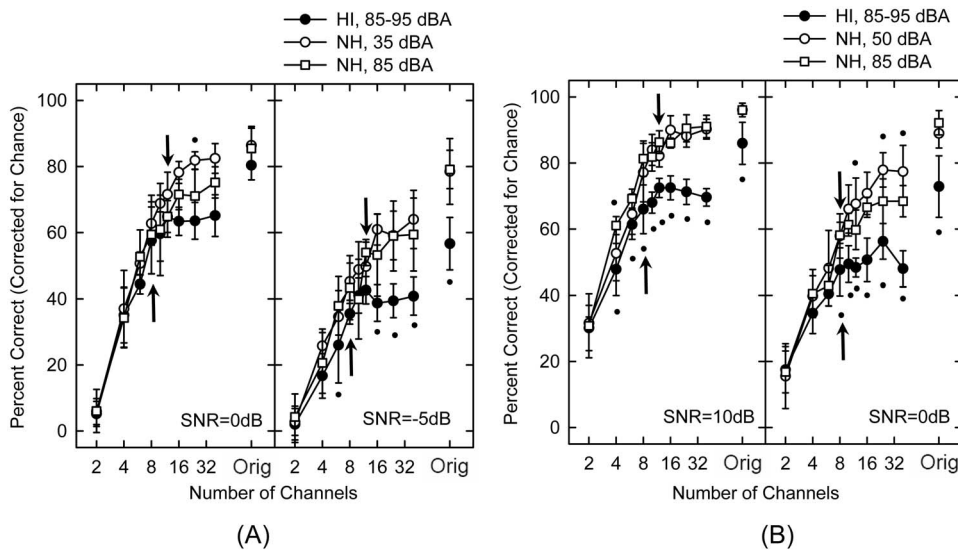
FIG. 6. Similar to Fig. 5, the filled and open circles show the scores by HI and NH listeners, respectively, replicated from experiment 1. The open squares show the percent correct scores by NH listeners with stimuli presented at the high level of 85 dBA.

significantly different than the scores by HI listeners (Table II). However, performance by NH subjects at this low level increased up to 12 channels; a larger number than 8 channels where performance by HI listeners reached the asymptote [Fig. 5(A); Table III].

In the conditions with processed consonants, the low presentation level did not change the performance by NH listeners at SNR=10 dB significantly; however, it resulted in a significant drop in scores at SNR=0 dB compared to the comfortable level, as shown by the significant main effect of level and the interaction ($p < 0.05$, Table V). Performance by NH listeners at the low level of 30 dBA was similar to performance by HI listeners for the SNR=10 dB noise level; but at SNR=0 dB there was a significant main effect of subject group ($p < 0.05$, Table II). The interaction between group and channels was significantly different at both noise levels ($p < 0.05$ at SNR=10 dB and $p < 0.01$ at SNR=0 dB, Table II), indicating a difference in the trends of the scores by NH and HI listeners. The Tukey test further showed that performance by NH subjects at the low presentation level increased up to 10 and 16 channels for the background noise conditions of SNR=10 dB and SNR=0 dB, respectively, while performance of HI subjects reached the asymptote at 8 channels [Fig. 5(B); Table III].

C. Experiment 3: Effects of high presentation level

In experiment 3, NH subjects were tested at a high presentation level of 85 dBA, comparable to the high presentation levels of 85–95 dBA used with HI listeners. The average percent correct scores at high levels, corrected for chance, are shown in Fig. 6 presented with scores from experiment 1. The open squares show the scores by NH listeners at the high presentation level of 85 dBA. The open circles show the scores by NH listeners from experiment 1, where vowels and consonants were presented at the comfortable levels of 35 and 50 dBA, respectively. The filled circles show the scores by HI listeners from experiment 1. The vowel recognition scores at SNR=0 dB and SNR=−5 dB are presented in the left and right panels, respectively, of Fig. 6(A). The consonant recognition scores at SNR=10 dB and SNR=0 dB are

presented in the left and right panels, respectively, of Fig. 6(B). The dots under the data show the conditions where the performance differed between NH and HI listeners, when both subject groups were tested at high presentation levels. The dots above the data show the conditions where the performance by NH listeners at comfortable and high presentation levels was different. The arrows below and above the data show the number of channels where the performance by HI and NH listeners at high levels, respectively, saturated.

In the unprocessed conditions, the scores by NH listeners did not change when the presentation level was increased from the comfortable levels of 35 and 50 dBA to the loud level of 85 dBA.

In the conditions with processed vowels, increasing the presentation level did not change the scores by NH listeners significantly; at both background noise conditions there was no significant main effect of level and no significant interaction (Table V). When performance at high presentation level was compared to performance by HI listeners, there was also no significant main effect of group and no interaction of group and channels at SNR=0 dB; however, performance by NH listeners saturated around 12 channels while performance by HI listeners saturated around 8 [Fig. 6(A), left panel; Table III]. There was a significant main effect of group and significant interaction at SNR=−5 dB ($p < 0.05$, Table II). The number of channels at saturation point was 12 for NH listeners, and 8 for HI listeners [Fig. 6(A), right panel; Table III].

In the conditions with processed consonants, increasing the presentation level did not change the consonant recognition scores by NH listeners significantly; however, there was significant interaction between level and number of channels in both background noise levels ($p < 0.05$, Table V). The scores by NH and HI listeners tested at similar absolute levels were significantly different in both background noise levels; there was a significant main effect of group and significant interaction ($p < 0.01$, Table II). At SNR=10 dB, performance by NH listeners asymptoted at 12 channels while performance by HI listeners asymptoted at 8 channels
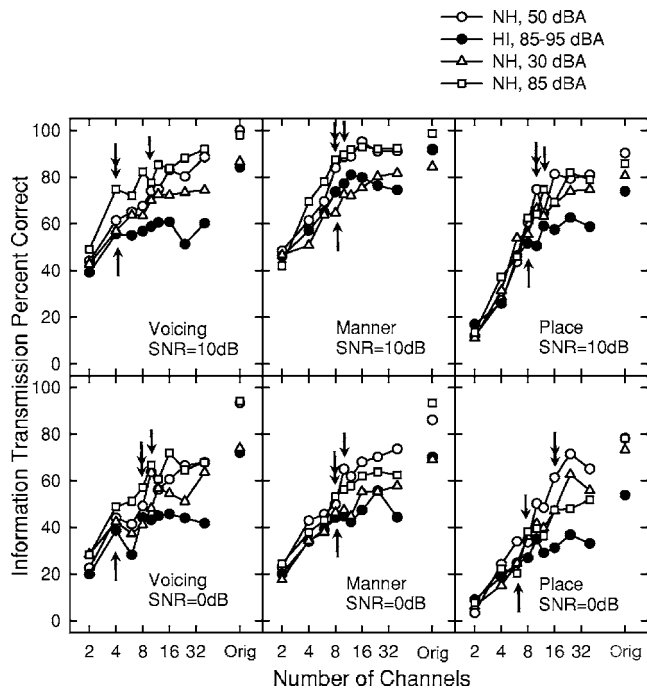
FIG. 7. Information transmission scores shown for the production based categories of voicing, manner, and place of articulation. The top and bottom rows present the scores obtained in the presence of background noise with SNR=10 dB and SNR=0 dB, respectively. The filled circles show the scores by HI subjects. The open symbols show the scores by NH subjects for three different presentation levels; open circles for 50 dBA, open triangles for 30 dBA, and open squares for 85 dBA. Arrows above the data show the number of channels at the asymptotic performance by NH listeners at different presentation levels. Arrows below the data show the number of channels at the asymptote for HI listeners.

[Fig. 6(B), left panel; Table III]. At SNR=0 dB, performance by both subject groups asymptoted at 8 channels [Fig. 6(B), right panel; Table III].

D. Analysis of consonant features

The ratio of spectral to temporal information required for the perception of the consonant features increases from voicing to manner to place of articulation (van Tasell *et al.*, 1992; Turner *et al.*, 1995). ter Keurs *et al.* (1992) and Boo-

throyd *et al.* (1996), for example, showed that spectral smearing affected the perception of the place feature most. To observe the effect of vocoder processing on perception of these features, consonant confusion matrices of experiments 1–3 were analyzed for the production based categories of voicing, manner, and place of articulation (Miller and Nicely, 1955).

The information transmission percent correct scores are presented for the noise conditions of SNR=10 dB and SNR=0 dB in the top and bottom panels of Fig. 7, respectively. The panels from left to right show the scores for voicing, manner, and place of articulation. Filled circles show the scores by HI subjects. Open symbols show the scores by NH subjects for three presentation levels; open circles for 50 dBA, open triangles for 30 dBA, and open squares for 85 dBA. The standard deviations were omitted for clarity.

The F values of the two-way mixed ANOVA that was used to compare the perception of features by NH and HI listeners are presented in Table VI. The F values of the two-way RM ANOVA that was used to compare the perception of features by NH listeners at different presentation levels are presented in Table VII. The number of channels where the performance reached the asymptote was determined by the Tukey test. These values are summarized in Table VIII, and are also shown in Fig. 7 by the arrows above and below the data for NH and HI listeners, respectively.

The left panels of Fig. 7 show that the voicing transmission patterns by NH listeners were similar at all presentation levels (Table VII) and performance by NH listeners was generally better than HI listeners (Table VI). The two-way mixed ANOVA showed that the interaction between group and channels was significantly different only in experiment 1 where the presentation level was set to the comfortable level of 50 dBA for NH listeners. The Tukey test confirmed this finding; performance by NH listeners reached the asymptote at 10 channels while performance by HI listeners reached the asymptote at 4 channels (Table VIII), in both background noise levels. In experiments 2 and 3, where the presentation levels for NH listeners were low (30 dBA) and high (85 dBA), respectively, both subject groups had the

TABLE VI. F values of the two-way mixed ANOVA, used to compare the performance by NH and HI subjects. The main effect of group and the interaction between group and channels are shown for production based categories of voicing, manner, and place of articulation for experiments 1–3.

		Voicing		Manner		Place	
	Noise	Group factor $F(1,7)$	Interaction $F(8,56)$	Group factor $F(1,7)$	Interaction $F(8,56)$	Group factor $F(1,7)$	Interaction $F(8,56)$
Expt 1: Comfortable level	SNR=10 dB	5.53[a]	2.75[a]	17.32[b]	2.06	46.59[b]	6.76[b]
	SNR=0 dB	3.58	2.34[a]	27.18[b]	4.98[b]	42.47[b]	7.36[b]
Expt 2: Low level	SNR=10 dB	2.25	0.84	0.21	1.74	2.13	1.77
	SNR=0 dB	9.62[a]	1.14	1.47	1.64	11.35[a]	4.09[b]
Expt 3: High level	SNR=10 dB	21.99[b]	1.89	59.39[b]	3.52[b]	20.34[b]	3.68[b]
	SNR=0 dB	27.90[b]	1.11	37.79[b]	1.90	12.13[b]	1.83

[a]$p<0.05$.
[b]$p<0.01$.

TABLE VII. F values of the two-way RM ANOVA, used to compare the performance by NH subjects at low and high presentation levels to the performance at comfortable level. The main effect of level and the interaction between level and channels are shown for production based categories of voicing, manner, and place of articulation for experiments 2 and 3.

| | Noise | Voicing | | Manner | | Place | |
		Level factor $F(1,4)$	Interaction $F(8,32)$	Level factor $F(1,4)$	Interaction $F(8,32)$	Level factor $F(1,4)$	Interaction $F(8,32)$
Expt 2:	SNR=10 dB	0.87	0.57	5.75	2.39[a]	0.32	1.38
Low level	SNR=0 dB	0.98	1.31	23.64[b]	1.95	11.48[a]	1.24
Expt 3:	SNR=10 dB	4.72	0.88	0.62	2.06	0.03	3.74[a]
High level	SNR=0 dB	0.93	0.57	3.00	1.62	12.23[a]	1.05[b]

[a] $p < 0.05$.
[b] $p < 0.01$.

asymptotic performance at 4 channels at SNR=10 dB noise level. At the higher noise level of SNR=0 dB, performance by NH subjects continued to increase up to 8–10 channels (Table VIII).

The middle panels of Fig. 7 show that the transmission of manner feature by NH listeners was similar at the levels of 50 and 85 dBA (Table VII), and at both levels NH listeners performed significantly better than HI listeners (Table VI). However, when the presentation level was reduced to 30 dBA, the performance dropped; the perception of manner by NH subjects at the low level of 30 dBA was similar to perception of manner by HI listeners (Table VI). The two-way mixed ANOVA showed a significant interaction of hearing loss and channels in experiment 1 for SNR=0 dB and experiment 3 for SNR=10 dB. However, the overall results of the Tukey test showed that the number of channels at performance saturation was similar (8 or 10) for both subject groups and in all experiments (Table VIII).

The right panels of Fig. 7 show that the transmission of place by NH listeners was similar at all presentation levels for SNR=10 dB. At SNR=0 dB, performance at the comfortable level of 50 dBA was significantly better than performance at the lower and higher presentation levels of 30 and 85 dBA (Table VII). There was also significant interaction between level and channels when the performance by NH listeners was compared at 50 and 85 dBA (Table VII). Performance by NH subjects was generally better than the HI

subjects, and there was a significant interaction between group and channels in most settings (Table VI). At SNR =10 dB, the number of channels where the performance saturated was slightly higher for NH listeners (10–12) compared to HI listeners (8; as shown in Table VIII). At SNR =0 dB, perception of place cue by HI listeners increased up to 6 channels. Place perception by NH listeners, on the other hand, increased up to 16 channels at the presentation levels of 30 and 50 dBA, and up to 8 channels at the presentation level of 85 dBA (Table VIII).

E. Experiment 4: Simulation of broad auditory filters

In experiment 4, the broad AFs were simulated with NH subjects by changing the filter slopes of the vocoder filters from a rollover of −24 dB\octave to a rollover of −6 dB\octave. In the first part, only the carrier filters were changed while the same analysis bands were used; this setup can be interpreted as a simulation of an input signal with good spectral resolution while the receiving component of the transmission system has reduced spectral resolution. In the second part, both carrier and analysis filters were made wider. This setup can be interpreted as a simulation of a transmission system with overall reduced spectral resolution.

The average vowel and consonant recognition scores, corrected for chance, are shown in Figs. 8(A) and 8(B), respectively, as a function of the number of spectral channels.

TABLE VIII. The number of channels where the performance by NH and HI listeners are saturated, shown for the production based categories of voicing, manner, and place of articulation.

| | Voicing | | Manner | | Place | |
	SNR=10 dB	SNR=0 dB	SNR=10 dB	SNR=0 dB	SNR=10 dB	SNR=0 dB
HI Expt 1 Comfortable level	4	4	8	8	8	6
NH Expt 1 Comfortable level	10	10	10	10	10	16
NH Expt 2 Low level	4	10	10	8	10	16
NH Expt 3 High level	4	8	8	8	12	8
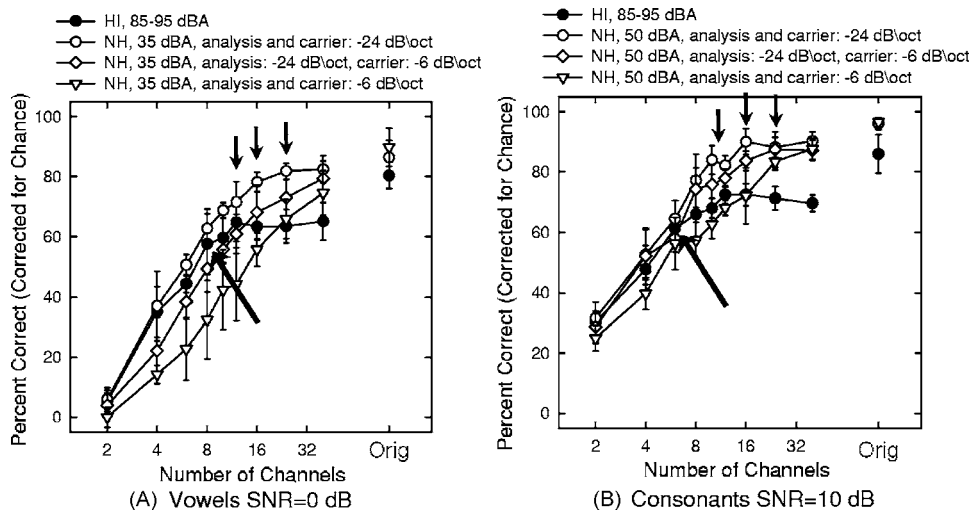
FIG. 8. Average phoneme recognition scores by NH listeners when tested with wide vocoder filters. The filled and open circles show the scores by HI and NH listeners, respectively, replicated from experiment 1. (A) and (B) show the vowel recognition scores with a background noise of SNR=0 dB and the consonant recognition scores with a background noise of SNR=10 dB, respectively. The open diamonds and the open triangles show the scores by NH listeners when the carrier filter slope was made shallower and when both the carrier and analysis filter slopes were made shallower, respectively. The arrows above the data show the number of channels where the performance by NH listeners reached the asymptote. The arrows below the data show the same for HI listeners.

The open and filled circles replicate the scores of HI and NH subjects, respectively, from experiment 1. The open diamonds show the average scores by NH subjects where only the carrier filter slopes were made shallower. The open triangles show the performance where both carrier and analysis filter slopes were made shallower. Vowels were presented in the background noise of SNR=0 dB, and consonants were presented in the background noise of SNR=10 dB. The presentation levels of the phonemes were the same comfortable level as in experiment 1; 35 dBA for vowels, and 50 dBA for consonants.

A two-way mixed ANOVA was used to compare the results by NH listeners tested with wide vocoder filters and the results by HI listeners. A two-way RM ANOVA was used

to compare the results by NH listeners with narrow and wide filters. The corresponding F values are presented in Table IX. The number of channels at the saturation point is shown in Table III, and is denoted by arrows above and below the data for NH and HI listeners, respectively, in Fig. 8.

When only the carrier filters were made wider, the performance by NH listeners did not change significantly (as shown by the open diamonds in Fig. 8 and also in the lower portion of Table IX). The number of channels where the performance saturated, however, increased from 10–12 to 16 (Table III). A stronger effect was observed when carrier and analysis filters were manipulated together (as shown by open triangles in Fig. 8). The scores by NH listeners dropped significantly and there was a significant interaction between fil-

TABLE IX. The upper portion; the F values of the two-way mixed ANOVA used to compare the performance by NH listeners with wide vocoder filters, from experiment 4, with the performance by HI listeners, from experiment 1. The lower portion; the F values of the two-way RM ANOVA used to compare the performance by NH listeners with narrow vocoder filters, from experiment 1, and wide vocoder filters, from experiment 4.

	Vowel recognition SNR=0 dB		Consonant recognition SNR=10 dB	
Two-way mixed ANOVA	Group factor $F(1,7)$	Interaction $F(8,56)$	Group factor $F(1,7)$	Interaction $F(8,56)$
Wide carrier filters	0.08	5.47	8.13[a]	6.41[b]
Wide carrier and analysis filters	11.28[a]	7.43[b]	0.12	9.02[b]

	Vowel recognition SNR=0 dB		Consonant recognition SNR=10 dB	
Two-way RM ANOVA	Filter slope factor $F(1,4)$	Interaction $F(8,32)$	Filter slope factor $F(1,4)$	Interaction $F(8,32)$
Wide carrier filters	2.87	2.55[a]	2.04	0.84
Wide carrier and analysis filters	25.52[b]	6.68[b]	29.45[b]	5.54[b]

[a]$p < 0.05$.
[b]$p < 0.01$.

ter bandwidth and channels (as shown in the lower portion of Table IX). The number of channels where the performance saturated increased from 10–12 to 24 (Table III).

Despite the substantial changes observed in the performance by NH listeners, either method failed to reproduce the trends seen in data by HI listeners. There was an overall reduction in the scores and the difference between the scores by NH and HI listeners was smaller as a result of these manipulations. However, the interaction between group and channels was generally significant (Table IX), indicating that the trends in the data by the two subject groups were different. The number of channels at the saturation point was higher with NH listeners than with HI listeners; 16 or 24 compared to 8 channels (Table III).

V. DISCUSSION

Turner *et al.* (1999) measured consonant recognition in quiet by NH and HI listeners as a function of the number of vocoder channels. The number of the channels varied from one to eight and the performance by HI listeners was poorer than NH listeners for all conditions, except for the single-channel processing. Yet, the performance by both subject groups increased as a function of the number of channels, contrary to the expectation that the performance by HI listeners would saturate at a smaller number of channels due to the limiting effect of reduced spectral resolution.

Friesen *et al.* (2001), however, observed a limiting effect of reduced spectral resolution on the perception of speech by CI users. When recognition of phonemes and sentences, processed with a noiseband vocoder, was measured with NH listeners as a function of the number of spectral channels, the performance increased up to 20 channels. Under similar listening conditions, the performance by CI users saturated around seven electrodes.

In the present study, it was hypothesized that a similar limiting effect of reduced spectral resolution might also be observed with HI listeners if the conditions from Turner *et al.* (1999) study were extended to a greater number of channels. Vowels were included in the stimuli as they were hypothesized to be more sensitive to spectral manipulations. The number of channels varied from 2 to 40, an upper limit selected on purpose higher than eight channels where the saturation in performance was observed with best CI users. Similar to the Friesen *et al.* (2001) study, background noise was added, as many studies showed that speech recognition by HI listeners in noise might be correlated with spectral resolution. Additionally, strict inclusion criteria were followed in the recruitment of the HI subjects and audibility was controlled by testing NH listeners at varying presentation levels.

In experiment 1, phoneme recognition was measured at similar sensation levels, determined from performance intensity functions by taking PI-50 levels as the reference for each subject group. Vowel recognition by NH and HI listeners, measured as a function of the number of vocoder channels, were similar in quiet and at the low background noise level of SNR=10 dB; the performance by both subject groups increased up to 8 channels before reaching plateau. Results

with consonant recognition in quiet were consistent with results by Turner *et al.* (1999); HI listeners generally had lower scores compared to NH listeners, but the trend in the performance was similar.

The limiting effects of reduced spectral resolution, similar to that reported by Friesen *et al.* (2001) with CI subjects, were observed with HI subjects only when the background noise level was increased. In all noise conditions, the performance by HI listeners saturated by 8 channels. Performance by NH listeners, on the other hand, increased up to 12 and 16 channels with vowel recognition at the noise levels of SNR =0 dB and SNR=−5 dB, respectively, and up to 10 and 12 channels with consonant recognition at SNR=10 dB and SNR=0 dB, respectively. This finding implies that while NH listeners could utilize a higher number of channels in challenging listening situations, such as when background noise was added, HI listeners did not show such ability.

The confusion matrices, obtained with consonants in background noise, were further analyzed for the transmission of the production based categories of voicing, manner, and place of articulation. Perception of the place feature is generally most sensitive to spectral degradations; therefore the effects were expected to be seen mainly with place. Consistent with this expectation, perception of place differed for NH and HI listeners; performance by HI listeners saturated by 6–8 channels while performance by NH listeners saturated by 10–16 channels. On the other hand, because perception of voicing and manner rely heavily on the temporal cues, the difference in the performance by NH and HI listeners was expected to be smaller with these features. Perception of manner was similar by the two subject groups. However, contrary to expectations, perception of voicing differed for NH and HI listeners; the number of channels at the saturation point was higher with NH listeners than with HI listeners when measured at similar SL (10 vs 4 channels).

A. Presentation levels

Audibility has a significant effect on speech recognition by HI listeners. It might become an even more important factor for challenging listening conditions such as perception of speech degraded by vocoder processing and presented in background noise. In experiment 2, audibility was reduced by testing NH subjects at quiet levels. When the levels were lowered by 15–20 dB compared to experiment 1, the percent correct scores by NH subjects dropped and the reduced performance was more comparable to the performance by HI subjects, for both processed and unprocessed conditions. Hence, audibility had a significant effect on speech recognition by NH subjects as well. However, even at these reduced performance levels, the trends in the performance by the two subject groups still differed. Performance by NH subjects increased up to 12 channels with vowels and up to 10–16 channels with consonants while performance by HI subjects saturated at 8 channels. This observation shows that the inability of the HI subjects to utilize more than eight spectral channels was probably not due to insufficient audibility.

Testing NH subjects at a relatively low presentation level, as it was done in experiment 1, is a reasonable attempt

to equalize the listening conditions for both subject groups in loudness, audibility, and comfort. However, by doing so, some potentially detrimental factors due to high presentation levels, such as broadening of AFs (Glasberg and Moore, 2000) or reduced speech recognition in noise (Studebaker *et al.*, 1999), might be overlooked.

To explore the effects of high presentation levels, NH subjects were tested at an absolute presentation level comparable to the presentation level used with the HI group (experiment 3). NH performance did not change significantly compared to experiment 1, for both processed and unprocessed conditions. The performance by NH subjects at the high presentation level was still significantly different when compared to HI listeners; the performance increased up to 12 channels with vowels and 8–12 channels with consonants. Therefore, it is unlikely that the effects observed in the study were due to distortions caused by high presentation levels. However, it is an interesting finding that no effect of AF widening, presumably due to high presentation levels, was observed on performance by NH listeners.

Note that in the present study the audiometric thresholds and the performance intensity functions were measured in quiet, and the presentation levels were adjusted based on these measurements. It is possible that the detection thresholds by HI listeners are considerably higher in noise, yet the present study did not control for this factor.

B. Spectral resolution

The noiseband vocoder was used to systematically degrade the speech stimulus. When background noise was added to the processing, it was observed that HI listeners were not able to use as many vocoder channels as NH listeners for the perception of phonemes. It was hypothesized that this effect was mostly due to reduced frequency resolution of hearing impairment. The inclusion criteria for HI subjects had been determined such that the probability of reduced spectral resolution was maximized; the subjects were selected from a group of patients with flat SNHL ranging from 50 to 60 dB HL, with no OAE and no apparent dead regions.

In most conditions with background noise, the number of channels where the performance by NH listeners saturated was 1.5–2 times larger than the number of channels where the performance by HI listeners saturated. With the assumption that the difference was mostly due to reduced spectral resolution of hearing impairment, this finding implies that the effective average spectral resolution by HI subjects was 1.5–2 times poorer compared to NH subjects in the noisy listening conditions. Even though it is not possible to deduce a conclusion about individual AF shapes for the HI subjects from this observation, it should be noted that this finding is within the ranges of AF broadening observed in SNHL in previous studies. Glasberg and Moore (1986) reported that AFs of impaired ears were broader compared to AFs in normal ears. When expressed in equivalent rectangular bandwidth (ERB), despite the large variation in the data, there were many subjects who showed a broadening in AF bandwidth by a factor of 2. Similarly, Stelmachowicz *et al.* (1985)

observed a reduction in Q10 values or the slopes of the low-frequency (LF) end of the psychophysical tuning curves (PTC), another measure for frequency resolution, by a factor of two for many HI subjects compared to NH subjects. These HI subjects had around 50 dB HL or more at the audiometric frequency of 2000 Hz, which is comparable to the hearing loss of the HI subjects in the present study.

Background noise was a major factor for observing the differences in the performance by NH and HI listeners. (1) The difference in the number of channels at the saturation point was observed only in noise. (2) There was a sharper drop in the performance by HI listeners in the unprocessed conditions as the noise level increased. (3) The additional drop in the performance from unprocessed to 40-channel processing condition was generally larger with HI listeners when there was background noise. These findings are consistent with earlier observations that speech perception by HI listeners in quiet is mainly determined by audibility (Festen and Plomp, 1983) while speech perception in noise is more sensitive to suprathreshold deficits (Noordhoek *et al.*, 2000), such as the reduced spectral resolution (Horst, 1987; van Schijndel *et al.*, 2001). In the present study, the difference between NH and HI listeners might have been small in quiet conditions because the audibility was maximized for HI listeners, and the effects of reduced spectral resolution might have showed in noise. One possible explanation for the reduced performance by HI listeners in noise was made by Leek and Summers (1996) who observed that HI listeners with broader AFs needed higher spectral contrast for recognition of speech in noise compared to NH listeners; they suggested that the wide AFs resulted in an internal representation with poorer SNR. The findings of the present study are consistent with this hypothesis; vowel recognition scores by HI listeners at SNR=10 dB, for example, were similar to scores by NH listeners at a higher noise level of SNR =0 dB.

In Experiment 4, the slopes of the vocoder bandpass filters were made shallower in an attempt to simulate broad AFs with NH listeners. Two simulation methods were used: (1) only the carrier filters were made wider, or (2) both the carrier and analysis filters were made wider. The first simulation method did not change the performance significantly. With the second method, the overall performance by NH subjects dropped and the scores were more comparable to the scores by HI listeners. However, the trends by the two subject groups were still different; the scores by NH listeners continued to increase to higher number of channels than HI listeners. Hence, the second simulation method also failed to reproduce the trend of the HI performance.

Fu and Nogaki (2005) used a similar spectral smearing method to simulate the reduced frequency selectivity of CI users with NH subjects. They were able to reproduce the performance by CI users in gated background noise, measured in Speech Reception Threshold (SRT), with NH listeners. However, they measured the SRT only with 4, 8, and 16 channel vocoders, as the main parameter of interest was the frequency of the gated noise. In the present study, the main interest was the change in overall performance as a function of the number of spectral channels, which varied over a

wider range of values (2–40) in steady background noise. Similar to Fu and Nogaki (2005), the performance by NH subjects dropped significantly with spectral smearing at 4, 8, and 16 channel conditions. However, when the performance by NH and HI listeners was compared over the entire range of the number of channels used, the trends were significantly different.

The failure to replicate the early saturating characteristic of the HI performance might indicate that the spectral sharpening mechanism of the healthy cochlea is highly effective, to the degree that NH listeners can make use of even minimal spectral contrast of heavily smeared speech stimulus. Baer and Moore (1993), for example, had to simulate AFs six times wider (in ERB) than normal to observe a degradation in speech recognition by NH listeners. A second possibility is that the method used in the present study to simulate broad AFs, where the same cutoff frequencies were used for the filters and only the filter slopes were made shallower, was not entirely adequate. As alternative methods, filter bandwidth could be increased to simulate broad critical bandwidth, as it was suggested by Dreschler and Plomp (1980), or asymmetrical filter shapes could be implemented, to simulate the irregular AF shapes observed with some HI listeners (Sommers and Humes, 1993).

C. Age effects

Some results of the present study were not entirely consistent with the assumption that the effects observed with the vocoder processing and in background noise were due to reduced spectral resolution of hearing impairment. For example, making the vocoder filter skirts shallower failed to replicate the performance by HI listeners with NH listeners, as it was discussed in the previous section. Also, when NH listeners were tested at a high level of 85 dBA, where AFs would presumably be wider (Glasberg and Moore, 2000), the saturation point of the NH performance did not move to a lower number of channels. Moore and Glasberg (2000) observed that the AF shape became asymmetrical with an elevation on the low-frequency side only. It is possible that the AF widening in NH listeners at high levels is to a smaller degree than the AF widening due to cochlear damage (Carney and Nelson, 1983). However, it is also possible that there were additional factors that affected the performance by HI listeners.

Previous research has shown that frequency resolution is generally related to the degree of hearing loss and is independent of age (Peters and Moore, 1992). A correlation with age is usually seen with temporal resolution; older listeners, for example, perform poorly in gap detection task (Snell, 1997; Strouse et al., 1998). This deficit due to aging is thought to be a possible factor for the difficulties that elderly people have in understanding speech in noise (Dubno et al., 1984; Snell and Frisina, 2000). In the present study, HI listeners were older than the NH listeners on average. This might have resulted in potentially different temporal processing abilities between the two subject groups and might have additionally affected the results.

Souza and Boike (2006) repeated the study by Turner et al. (1999) with HI subjects of varying age. The performance by older subjects was generally lower compared to younger subjects for all vocoder processing conditions. However, the trend in data was similar; recognition of consonants increased up to 8 channels by all subjects. When the consonant features were analyzed for voicing, manner and place, the only feature that was correlated with age was voicing; older subjects had similar perception of manner and place as younger subjects, but the perception of voicing was poorer by the older subjects.

In the present study, the place cue was expected to be affected most by the spectral degradations. However, in addition to place cue, the transmission of voicing was also affected. The similarity of this finding to observations by Souza and Boike (2006) implies that the HI subjects of the present study might have had deficits in both temporal and spectral processing mechanisms, and the effects of these deficits might have been observed in the perception of different features.

An interesting extension to the present study would be to acquire individual measures from subjects for spectral and temporal resolution, and to correlate these measures with performance by a specific subject. Such a study would provide a more definitive answer to the question if the effects observed in the present study were mainly due to reduced spectral resolution as it was hypothesized.

D. Comparison with implant users

Friesen et al. (2001) observed that performance by CI users saturated at a smaller number of channels compared to NH listeners. To explore how the results with implant users compare to results with HI listeners, vowel recognition scores were reproduced from Friesen et al. (2001) and superimposed with scores from the present study, for corresponding noise conditions. In Fig. 9, the open and filled circles show the vowel recognition scores from experiment 1 for NH and HI listeners, respectively. The hatched area shows the range of scores, with upper and lower borders defined by best and worst performance by CI users, respectively. The solid line shows the average performance by Nucleus-22 implant users. The scores were measured using the same vowel stimuli in both studies, and the scores with implant users were also corrected for chance level for consistency.

The combined results in Fig. 9 show that at SNR =0 dB, the performance by CI and HI listeners similarly saturated at smaller number of channels (i.e., 7 and 8, respectively) compared to NH listeners (12 channels). The average CI performance was considerably lower than average NH and HI performance, but the best CI performance was similar to the average HI performance. These observations are consistent with results reported by Henry et al. (2005) that showed that the frequency resolution with HI listeners was poorer than NH listeners and better than CI listeners, and only the best implant users performed at levels similar to HI listeners.

The poor peformance by implant users, compared to NH and HI listeners, show that there are probably many factors that affect speech recognition by CI users. After all, electric
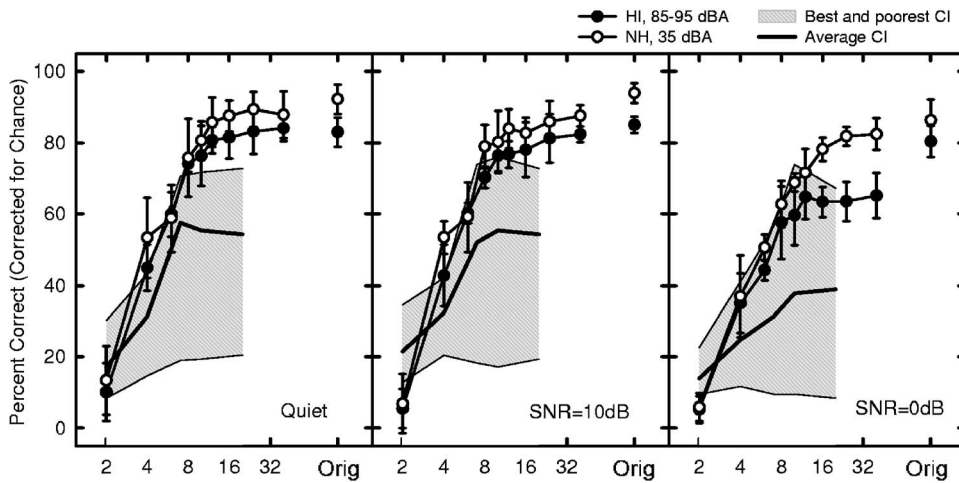
FIG. 9. Average vowel recognition scores by NH and HI listeners presented with average scores by CI users replicated from the study by Friesen *et al.* (2001). From left to right panels the background noise increases. The open and filled symbols show the performance by NH and HI listeners, respectively. The solid line shows the average scores by Nucleus-22 implant users. The hatched area shows the range of the scores by CI subjects.

hearing works on an entirely different mechanism than the acoustic hearing. On the other hand, there was at least one setting (SNR=0 dB) of the present study that produced similar trend in performance by HI and CI listeners. This similarity can be interpreted that an auditory system with inherently reduced frequency resolution, either due to a loss in the peripheral nonlinear mechanism or maybe even due to deficits in the central auditory system, might have a limiting effect on performance similar to CI users whose performance is believed to be limited mainly due to channel interactions. On a more speculative note, one can hypothesize that there might be such factors in addition to channel interactions that are further degrading the CI performance. However, the results of the current study did not completely support this hypothesis as the effects with HI listeners were seen only in noise and there were indications that the performance by HI listeners might have been additionally affected by factors such as reduced temporal resolution, which is usually not observed with implant users.

ACKNOWLEDGMENTS

American National Standards Institute (**1997**). Methods for Calculation of the Speech Intelligibility Index, S35.

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Baker, R. J., and Rosen, S. (**2002**). "Auditory filter nonlinearity in mild/moderate hearing impairment," J. Acoust. Soc. Am. **111**, 1330–1339.

Bierer, J. A., and Middlebrooks, J. C. (**2002**). "Auditory cortical images of cochlear-implant stimuli: Dependence on electrode configuration," J. Neurophysiol. **87**, 478–492.

Bierer, J. A., and Middlebrooks, J. C. (**2004**). "Cortical responses to cochlear implant stimulation: Channel interactions," J. Assoc. Res. Otolaryngol. **5**, 32–48.

Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (**1996**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807–1818.

Carney, A. E., and Nelson, D. A. (**1983**). "An analysis of psychophysical tuning curves in normal and pathological ears," J. Acoust. Soc. Am. **73**, 268–278.

Cazals, Y., Pelizzone, M., Kasper, A., and Montandon, P. (**1990**). "Multichannel cochlear implant patients with different open speech understanding show some similar basic psychophysical results," Acta Oto-Laryngol. **S469**, 150–155.

Chatterjee, M., and Shannon, R. V. (**1998**). "Forward masked excitation patterns in multielectrode cochlear implants." J. Acoust. Soc. Am. **103**, 2565–2572.

Dreschler, W. A., and Plomp, R. (**1980**). "Relation between psychopchysical data and speech perception for hearing-impaired subjects," J. Acoust. Soc. Am. **68**, 1608–1615.

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). "Effects of age and mild hearing loss on speech recognition in noise," J. Acoust. Soc. Am. **76**, 87–96.

Festen, J. M., and Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am. **73**, 652–662.

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," J. Speech Lang. Hear. Res. **40**, 1201–1215.

Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Garnham, C., O'Driscoll, M., Ramsden, R., and Saeed, S. (**2002**). "Speech understanding in noise with a Med-El 40+ cochlear implant using reduced channel sets," Ear Hear. **23**, 540–552.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.

Gstoettner, W. K., Adunka, O., Franz, P., Hamzavi, J., Plenk, H., Susani, M., Baumgartner, W., and Kiefer, J. (**2001**). "Perimodiolar electrodes in cochlear implant surgery," Acta Oto-Laryngol. **121**, 216–219.

Henry, B., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal-hearing, hearing-impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hornsby, B. W. Y., Trine, T. D., and Ohde, R. N. (**2005**). "The effects of high presentation levels on consonant feature transmission," J. Acoust. Soc. Am. **118**, 1719–1729.

Horst, J. W. (**1987**). "Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects," J. Acoust. Soc. Am. **82**, 874–885.

Kwon, B. J., and van den Honert, C. (**2006**). "Effect of electrode configu-

Deniz Başkent: Frequency selectivity in sensorineural hearing loss

ration on psychophysical forward masking in cochlear implant listeners," J. Acoust. Soc. Am. **119**, 2994–3002.

Leek, M., and Summers, V. (**1996**). "Reduced frequency selectivity and the preservation of spectral contrast in noise," J. Acoust. Soc. Am. **100**, 1796–1806.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Moore, B. C. J. (**1996**). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," Ear Hear. **17**, 133–160.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). "A test for diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2000**). "Measuring the threshold for speech reception by adaptive variation of the signal bandwith. II. Hearing-impaired listeners," J. Acoust. Soc. Am. **107**, 1685–1696.

Norton, S. J. (**1992**). "Cochlear function and otoacoustic emissions," Semin. Hear. **13**, 1–14.

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Peters, R. W., and Moore, B. C. J. (**1992**). "Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects," J. Acoust. Soc. Am. **91**, 256–266.

Pfingst, B. E., Zwolan, T. A., and Holloway, L. A. (**1997**). "Effects of stimulus configuration on psychophysical operating levels and on speech recognition with cochlear implants," Hear. Res. **112**, 247–260.

Pfingst, B. E., Franck, K. H., Xu, L., Bauer, E. M., and Zwolan, T. A. (**2001**). "Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses," J. Assoc. Res. Otolaryngol. **2**, 87–103.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.

Shannon, R. V., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.

Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.

Snell, K. B., and Frisina, D. R. (**2000**). "Relationships among age-related differences in gap detection and word recognition," J. Acoust. Soc. Am. **107**, 1615–1626.

Sommers, M. S., and Humes, L. E. (**1993**). "Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners," J. Acoust. Soc. Am. **93**, 2903–2914.

Souza, P. E., and Boike, K. T. (**2006**). "Combining temporal-envelope cues across channels: effects of age and hearing loss," J. Speech Lang. Hear. Res. **49**, 138–149.

Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., and Mott, J. (**1985**). "Speech perception ability and psychophysical tuning curves in hearing impaired listeners," J. Acoust. Soc. Am. **77**, 620–627.

Stickney, G. S., Loizou, P. C., Mishra, L. N., Assman, P. F., Shannon, R. V., and Opie, J. M. (**2006**). "Effects of electrode design and configuration on channel interactions," Hear. Res. **211**, 33–45.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelope smearing on speech reception. I," J. Acoust. Soc. Am. **91**, 2872–2880.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Effect of spectral envelope smearing on speech reception. II," J. Acoust. Soc. Am. **93**, 1547–1552.

Throckmorton, C. S., and Collins, L. M. (**1999**). "Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear implant subjects," J. Acoust. Soc. Am. **105**, 861–873.

Turner, C. W., and Henn, C. C. (**1989**). "The relation between vowel recognition and measures of frequency resolution," J. Speech Hear. Res. **32**, 49–58.

Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.

Turner, C. W., Chi, S.-L., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **42**, 773–784.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**2001**). "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 529–542.

van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). "Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants," J. Acoust. Soc. Am. **92**, 1247–1257.

van Tasell, D. J. (**1993**). "Hearing loss, speech, and hearing aids," J. Speech Hear. Res. **36**, 228–244.

Wilson, B., Finley, C., Laeson, D., Wolford, R., Eddington, D., and Rabinowitz, W. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.

Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.

Auditory measures of selective and divided attention in young and older adults using single-talker competition

Larry E. Humes, Jae Hee Lee, and Maureen P. Coughlin
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405-7002

In this study, two experiments were conducted on auditory selective and divided attention in which the listening task involved the identification of words in sentences spoken by one talker while a second talker produced a very similar competing sentence. Ten young normal-hearing (YNH) and 13 elderly hearing-impaired (EHI) listeners participated in each experiment. The type of attention cue used was the main difference between experiments. Across both experiments, several consistent trends were observed. First, in eight of the nine divided-attention tasks across both experiments, the EHI subjects performed significantly worse than the YNH subjects. By comparison, significant differences in performance between age groups were only observed on three of the nine selective-attention tasks. Finally, there were consistent individual differences in performance across both experiments. Correlational analyses performed on the data from the 13 older adults suggested that the individual differences in performance were associated with individual differences in memory (digit span). Among the elderly, differences in age or differences in hearing loss did not contribute to the individual differences observed in either experiment. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354070]

I. INTRODUCTION

A common complaint of older adults is that they can hear speech, but cannot understand it. This is especially true when listening to speech in a background of noise or while other people are talking. A working group of the Committee on Hearing and Bioacoustics and Biomechanics (CHABA) of the National Research Council, reviewing the work on the speech communication problems of older adults available through the early 1980s, offered several hypotheses regarding the causes of such communication problems (CHABA, 1988). Briefly, these hypotheses were (1) the peripheral auditory hypothesis, which suggested that the difficulties were primarily attributable to the sensorineural hearing loss and cochlear pathology that is common among older adults; (2) the central-auditory hypothesis, which posited age-related, modality-specific, changes in the auditory portions of the central nervous system from the lower brainstem through auditory centers of the cortex; and (3) the cognitive hypothesis, which suggested age-related decline in general cognitive functions, such as memory, attention, and speed of processing. The working group also recognized that it was possible to have various combinations of these factors at work in a given individual.

In the years since this CHABA working group report, our laboratory, among others, has demonstrated repeatedly that hearing loss, or audible bandwidth, represents the primary (but not sole) factor underlying the problems experienced by older adults when listening to speech in quiet or in noise (van Rooij, Plomp and Orlebeke, 1989; Jerger, Jerger, and Pirozzolo, 1989; van Rooij and Plomp, 1990, 1992; Helfer and Wilber, 1990; Humes and Roberts, 1990; Humes and Christopherson, 1991; Lee and Humes, 1993; Humes *et al.*, 1994; Souza and Turner, 1994; Humes, 1996; Divenyi

and Haupt, 1997a, 1997b; Humes, 2002). Most recently, we found this to be true even for time-compressed speech presented at high sound levels designed to reduce, but not eliminate, the negative impact of high-frequency sensorineural hearing loss (Humes, 2005).

For most older adults with speech-communication problems, the best way to address these problems is through amplification, such as that provided by personal hearing aids. Hearing aids fit to older individuals following typical clinical procedures, however, do not always guarantee the full audibility of speech to the wearer (Humes, 1991; Humes and Halling, 1993). In that regard, it is not surprising that other studies, including another study from our laboratory, examining *aided* speech communication in quiet and noise in older adults and using hearing aids fit under typical clinical procedures, still found that audibility of the speech stimuli played a primary role in determining individual differences in performance (Humes, 2002; Walden and Walden, 2004).

In more recent work in our laboratory with older adults, we have used laboratory-based approaches to amplification to assure suitable audibility of the speech stimuli through at least 4000 Hz. The first such study in this series (Coughlin, 2004) examined the speech-identification differences between young normal-hearing adults and older hearing-impaired listeners when the Coordinate Response Measure (CRM; Bolia *et al.*, 2000) was used as both the target and competing speech, with separate talkers used for each. The CRM is a sentence-length test that requires the listener to identify the last two words of the sentence from a closed set of alternatives (one of four colors followed by one of eight numbers; e.g., "red 4", "blue 2"). When audibility through 4000 Hz was assured, hearing loss no longer explained individual differences in speech-identification performance among the 19 older adults in that experiment. Yet, when the

performance of young adults was compared to older adults, the latter still needed about a 6-dB better target-to-competition ratio than the young adults to achieve the same levels of performance under otherwise equivalent conditions. Thus, it would appear that, when audibility has been eliminated as a factor, other factors may emerge that impact differences in performance between younger and older adults or among older adults.

In keeping with the conceptual framework laid out by the CHABA working group (CHABA, 1988), other possibilities could include age-related modality-specific changes in central-auditory processing or amodal cognitive processing. In the CRM speech-identification task used by Coughlin (2004), two similar sentences were presented simultaneously to the same ear and the target message was identified by a name (or "call sign") at the beginning of the target sentence. Thus, the listeners needed to attend to the target sentence associated with a particular call-sign ("Baron") and ignore a very similar competing message spoken by a different talker and using a different call sign (e.g., "Charlie," "Laker," etc.). As a result, the task required the listener to selectively attend to the target sentence identified by the target call sign. Consequently, it is possible that age-related differences in selective attention could have contributed to the older adults' need for a better target-to-competition ratio than required by the young adults. Age-related differences in selective attention have been observed previously [see reviews by McDowd and Shaw (2000) and Rogers (2000)], although, for at least some measures of selective attention, these age effects have been attributed to memory deficits or age-related mental slowing (Verhaeghen and De Meersman, 1998a, 1998b) rather than attention deficits *per se*.

In the present study we sought to examine age-related differences in auditory measures of attention using the same CRM single-talker-competition paradigm used by Coughlin (2004). Auditory measures of both selective and divided attention were included in this study to provide a broader picture of possible age-related changes in attention. Age-related deficits in divided attention have also been reported previously (see the review by McDowd and Shaw, 2000), although general age-related mental slowing may again play a role in such deficits (Verhaeghen *et al.*, 2003).

Another issue examined in this study had to do with the relative strength of various auditory cues in selective and divided attention. In particular, does it matter whether the cue for the auditory measures of attention is one based on localization/lateralization (ear), lexical information (call sign), or acoustical/indexical information in the speech stimulus (talker gender)? In Experiment 1, monaural and dichotic presentation conditions were used to measure auditory selective and divided attention with the cues being either ear or call sign. The match between the gender of the target and competing talkers was also manipulated as a repeated-measures independent variable in this experiment. In general, for young normal-hearing listeners, the CRM speech-identification task is easier when the gender of the target and competing talker differ, presumably due to better segregation of the two auditory objects (Brungart, 2001; Coughlin, 2004). It was hypothesized that those cues associated with

physical aspects of the perceived auditory object, such as its location (ear) or fundamental frequency (gender), will prove stronger than the arbitrary lexical labels (call signs) attached to those objects.

In experiment 2, the effect of talker uncertainty on measures of attention was also examined, with the range of conditions varying from the same male and female talkers for a given block of trials to the specific talker randomly selected from a set of four for each gender from trial to trial. Brungart and Simpson (2004) have previously reported little effect of talker uncertainty on the CRM task for young adults, whether the testing was conducted monauralally or dichotically. In other contexts, however, older adults have been found to show larger decrements in performance with increasing talker uncertainty (Sommers, 1997). In addition, Coughlin (2004) demonstrated the effects of talker uncertainty with the CRM test in both young normal-hearing and elderly hearing-impaired adults. As a result, the effects of talker uncertainty were examined further in experiment 2.

Finally, as noted, one of the important procedural aspects of the earlier study by Coughlin (2004) was the use of spectrally shaped speech materials for the older adults with hearing loss. In particular, in that study spectral shaping was applied to ensure audibility of the CRM speech materials through 4000 Hz. By doing so, there was no significant correlation between performance on the CRM and high-frequency hearing thresholds. Steps were taken in this study as well to ensure sufficient audibility of the CRM speech stimulus through at least 4000 Hz for all listeners. Details are provided in the next section.

II. METHODS

A. Participants

Two groups of listeners differing in age and hearing loss were tested: (1) elderly hearing impaired (EHI) and (2) young normal hearing (YNH). There were 10 YNH participants, 5 males and 5 females, for experiment 1 and 2 with 7 of the 10 who participated in experiment 1 also participating in experiment 2. The mean age for the YNH participants in experiment 1 was 25.4 years and ranged from 21 to 34 years whereas the mean age was 25.9 years and ranged from 22 to 34 years for those in experiment 2. All YNH subjects passed a pure-tone air-conduction screening at 20 dB HL (ANSI, 1996) from 250 through 8000 Hz in octave intervals, passed the Mini-Mental Status Exam (a MMSE score of at least 27 of 30; Folstein, Folstein, and McHugh, 1975), and had forward digit spans of at least 5 (i.e., could recall sequences of at least 5 digits) and backward digit spans of at least 4.

There were 13 EHI participants, 9 male and 4 female, and the same 13 completed both experiments in this study. These individuals ranged in age from 61 to 81 years with a mean age of 76.2 years. The EHI listeners had varying degrees of sensorineural hearing loss, as illustrated in Fig. 1. Three EHI listeners had interaural asymmetries in pure-tone thresholds exceeding 15 dB for at least one frequency, but all such asymmetries were at frequencies of 4000 Hz or higher.
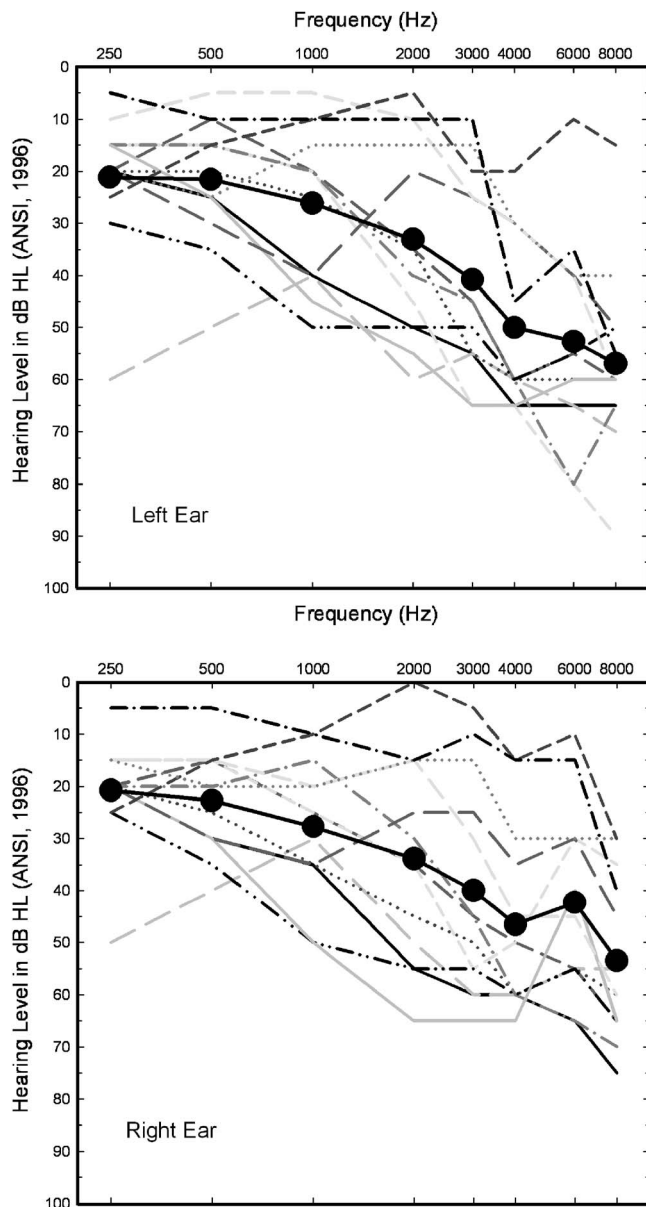
FIG. 1. Air-conduction hearing thresholds for the left (top) and right (bottom) ears of each of the 13 elderly hearing-impaired (EHI) listeners in this study. The same line types are used to represent the audiograms of each subject in each panel. The filled circles depict the mean hearing thresholds for each ear.

All 13 participants passed the MMSE (a score of at least 27) and had forward digit spans of at least 5 and backward digit spans of at least 4.

All EHI subjects were community dwelling at the time of testing. They were paid $10/hour for their participation. The YNH participants were recruited from the general Indiana University (IU) campus community via flyers and the undergraduate and graduate divisions of the IU Department of Speech and Hearing Sciences via e-mail. They were paid $7/h for their participation.

B. Stimuli and apparatus

The Coordinate Response Measure corpus (CRM; Bolia *et al.*, 2000) was chosen as a sentence-level speech test for this project. The structure of each sentence is identical, with

each sentence using the form "Ready (call sign), go to (color) (number) now." By instructing the listener to attend only to the designated call sign while ignoring competing sentences, it is possible to investigate the effectiveness of masking or interference from competing speech signals. The corpus consists of every combination of the eight call signs (arrow, baron, charlie, eagle, hopper, laker, ringo, tiger), four colors (blue, green, red, white) and eight numbers (1–8), resulting in 256 phrases for each talker. Four male and four female talkers yield a total of 2,048 phrases or sentences. Brungart (2001) reported that two talkers, one of each gender, yield scores when used as a target or competing talker that differ somewhat from the remaining talkers. Therefore, in experiment 1, only six talkers (three males and three females) were used. In experiment 2, in which talker uncertainty was manipulated, it was decided that all eight talkers should be used to expand the amount of talker uncertainty available.

Using CoolEdit Pro 2000 (Syntrillium, version 1.2a), the CRM sentences were converted from binary files (*.bin) to wave files (*.wav) and from stereo to mono recordings for use in a custom MATLAB (Mathworks, version 6.5) program. The MATLAB program was designed to present two stimuli simultaneously through Tucker Davis Technologies (TDT) System-III equipment. The sampling rate was 48 828 Hz. The stimuli were either added digitally and routed through one channel of a 16-bit digital-to-analog converter (DA1) for monaural presentation or presented separately through channel 1 (target) and 2 (competition) of the DA1 for dichotic presentation. The stimuli were then sent to a headphone buffer (TDT HB7) and finally output via matched insert earphones (E-A-R, ER-3A).

1. Calibration

The CRM corpus was designed so that each sentence was equated for the total rms power. For a calibration of the unshaped CRM used with the YNH listeners, a 30 s calibration noise and calibration tone (1000 Hz) were created and equated using Cool Edit Pro 2000. The calibration tone was matched to the peak levels for the speech materials and was used to determine the maximum possible presentation level without peak clipping. Once this maximum output setting was established, the calibration noise was used to measure the maximum levels in one-third octave bands from 125 through 8000 Hz.

The calibration noise was matched in spectrum and average rms amplitude to the long-term spectrum and average rms amplitude of the CRM materials. To determine the long-term average speech spectrum values of the CRM stimuli, the target CRM sentences used in this study were digitized using Cool Edit Pro 2000 and concatenated to create one large file of all eight talkers. The average amplitude spectrum of the concatenated wave file was then measured and the calibration noise was shaped digitally using a 1/3-octave-band graphic equalizer within Cool Edit Pro 2000 until the shaped noise was within 3 dB of the long-term-average speech spectrum for the CRM materials from 100 through 7500 Hz. This calibration stimulus was then presented through the TDT hardware and insert earphones
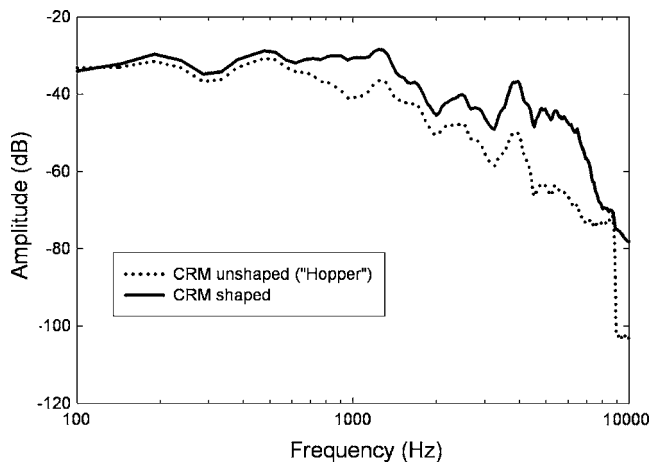
FIG. 2. The relative long-term average rms amplitude spectrum (50 ms window) for all CRM sentences using one of the eight call signs ("Hopper") before spectral shaping (dotted line) and after spectral shaping (solid line). This figure illustrates the amount of spectral shaping applied to all CRM materials for the 13 EHI listeners.

coupled to an HA-2 2 cm³ coupler (ANSI, 1996). Overall (linear) and one-third-octave band sound pressure levels were measured using a Larson Davis model 800B sound level meter (with a Larson Davis model 2575 1 in. microphone) and recorded. An overall sound pressure level of 87 dB SPL for the unshaped CRM stimuli was used with all YNH listeners.

Next, given the presence of varying degrees of sloping, high-frequency sensorineural hearing loss in all the older listeners (see Fig. 1), the CRM materials underwent spectral shaping to increase the amplitude at 125, 250, 500, 1000, 2000, 4000, and 6000 Hz by 2, 2, 2, 5.5, 10.5, 11.5, and 20 dB, respectively. This effectively flattened the one-third-octave-band spectrum acoustically at 100 dB SPL (±2 dB) from 200 through 4000 Hz in the HA-2 2 cm³ coupler. A comparison of the amplitude spectra for the shaped and unshaped stimuli is provided in Fig. 2 for a sample of CRM sentences (those making use of the call sign, "Hopper"). The digital wave files for the spectrally shaped CRM sentences were examined to find the peak amplitude following shaping, which turned out to be the color "red" spoken by one of the talkers. This segment of the waveform was excised, copied, and concatenated to produce a long calibration file that was used to set the peak amplitude of the playback system to avoid peak clipping. The same spectral shaping was then applied to the calibration noise and the overall sound pressure level and 1/3-octave-band levels in the HA-2 2 cm³ coupler were measured for this calibration noise using a Larson Davis model 800B sound level meter (with a Larson Davis model 2575 1 in. microphone) and recorded. The maximum sound pressure level, with no attenuation in the system and without peak clipping, was 114 dB SPL. With full attenuation at the headphone buffer, the presentation level was 87 dB SPL. The 1/3-octave-band level at a center frequency of 4000 Hz (73 dB SPL) was then noted for this default presentation level. As long as the listener's pure-tone hearing threshold at 4000 Hz in the right ear, expressed in dB SPL *re*: 2 cm³ coupler, was at least 15 dB lower (⩽58 dB SPL) than 1/3-octave-band level of the CRM calibration

noise at a center frequency of 4000 Hz, then the default overall presentation level of 87 dB SPL was used. This ensured that the rms spectrum of the speech stimulus would be at least 15 dB above the hearing threshold through 4000 Hz. The assumption here is that hearing threshold for pure tones and 1/3-octave-band noises are equivalent in hearing-impaired listeners (Cox and McDaniel, 1986). If the listener's hearing threshold at 4000 Hz was such that the 1/3-octave-band level of the CRM stimuli at that same frequency was not at least 15 dB above threshold, then the attenuation of the headphone buffer was decreased in 3-dB steps until this goal was achieved. Given the headphone-buffer step size of 3 dB, this means that the speech energy at 4000 Hz was 15–17 dB above threshold at that frequency. Given the speech spectra (Fig. 2) and the patterns of hearing loss (Fig. 1), this also ensured that the long-term-average rms speech spectrum was at least 15 dB above threshold at lower frequencies as well. All of the YNH subjects and 5 of the 13 EHI subjects were tested at the default level of 87 dB SPL, although, as noted, the YNH listeners did not listen to the spectrally shaped CRM materials. For the remaining 8 EHI listeners, 2 were tested at 90 dB SPL, 5 at 99 dB SPL, and 1 at 105 dB SPL. For all conditions in this study, the target and competing sentences were presented at the same sound pressure level for a 0 dB target-to-competition ratio (TCR).

C. General procedures

After administering the initial MMSE and digit-span measures, similar procedures were followed in experiments 1 and 2. The procedural details in common to both experiments are reviewed here.

The subject was seated in a single-walled sound booth in front of a computer monitor and was given written instructions for the CRM test. CRM test presentation was designed so that participants would hear 32 presentations of competing sentence pairs per test block. Thirty-two presentations per test block allowed for representation of each color and number combination as the target without repetition. The participants first received one 32-sentence block as practice in each cue condition. The participants listened to a total of 4 blocks of 32 sentence pairs, or 128 sentence pairs per condition. The listener indicated the color and number coordinates of the target sentence by selecting one of four rectangular buttons arranged in a column on the left side of the computer screen, labeled by both the color name and the color itself (i.e., a red rectangle with a text label of "red"), and then selecting a number from a column of numbers listing 1 through 8 on the right side of the computer screen. Most subjects in both groups used a computer mouse to make their color-number selections, but if any subjects, particularly among the EHI group, had difficulty with the use of the computer mouse, a touch-screen monitor was used instead. Trial-to-trial feedback was not provided.

In both experiments, a large-font (40-point, Times New Roman) orthographic cue (call sign, right/left, or male/female), which signaled the target sentence for that particular block of trials, was presented three times in the upper left portion of the computer screen. To minimize confusion with

other colors displayed simultaneously on the screen, the cue was displayed in a purple font, centered within a white rectangular box measuring 1.5 cm vertically and 4.5 cm horizontally. The duration of the orthographic cue was 0.8 second, with an interval of 0.2 second between each presentation. Thus, the orthographic cueing of the target sentence required a total duration of 3 s. For measures of selective attention, on each trial in a block, the 3 s cue interval ended immediately before the presentation of a sentence pair. For selective attention, the listener could respond by selecting the target color-number coordinates immediately after the presentation of the sentence pair had been completed. For divided attention, the 3 s cue interval was initiated immediately following the presentation of the sentence pair and responses were not possible until the end of the cue interval. Because of the different memory constraints imposed on the measures of selective and divided attention in this study, these two measures are considered to be separate dependent variables in subsequent analyses (rather than two values of a single independent variable, attention type).

When the cue was ear (experiment 1), the words "Right" or "Left" were used as the orthographic cue. When the cue was call sign (experiment 1), the cue was one of the eight possible call signs for the CRM (e.g., "Charlie"). When the cue was talker gender (experiment 2), the words "Male" or "Female" were used as the orthographic cue. In each block of 32 trials for a given cue condition, each possible cue value occurred an equal number of times and the order was randomized. For example, when the call sign was the cue, in a given block of 32 trials, each of the 8 possible call signs appeared as a target four times, in random order, and competing call sign values were also equally distributed in random order (with the exception that the same call sign could not be used for both target and competing sentences on a given trial).

The EHI listeners also received a control condition in which two blocks of 32 trials of the CRM were presented to each ear without a competing stimulus. This was used to verify that these listeners could, in fact, identify the color-number coordinates with a high degree of accuracy given the spectral shaping provided.

The CRM testing was self-paced with the inter-trial interval controlled by the subject. The elderly subjects showed some variability in the time required to complete the testing. In general, with each session requiring 90–120 min to complete, older listeners required more sessions (6–8) than the younger listeners (5) to complete all testing. Additional experiment-specific procedural details are provided below.

III. EXPERIMENT 1

A. Experiment-specific procedures

There were a total of six conditions examined in this experiment for selective attention and six for divided attention. For both selective and divided attention, two of six conditions made use of dichotic stimulus presentation and the ear cue, two used dichotic presentation and the call-sign cue, and two used monaural presentation and the call-sign cue. An additional independent variable examined in experiment 1 was the agreement between talker genders for the target and competing speech stimuli. Target and competing talkers were either the same or different gender for a given block of trials. When the talkers were the same gender, each was chosen randomly without replacement from among a pool of 3 talkers on each trial. When the talkers were of opposite gender, each was chosen randomly from a pool of 3 talkers for each gender on each trial. For target and competing talkers of either the same or different gender, random selection was constrained so that male and female talkers were each used as the target talker half the time. For all dichotic stimulus presentation conditions, half the trials in a block presented the target stimulus to the right ear and half to the left ear in random order. For monaural presentation, the right ear was always the stimulus ear. As noted previously, three of the elderly subjects had an interaural asymmetry of pure-tone thresholds exceeding 15 dB, but these were all confined to frequencies of 4000 Hz or higher, and the right ear (test ear) always had the better hearing thresholds.

The six conditions for each type of attention measure in this experiment represent an unbalanced research design because it is not possible to use the ear cue in monaural listening. As a result, there are two overlapping sets of 2×2 factorial repeated-measures combinations represented in the six conditions. When the dichotic ear-cue conditions are compared to the dichotic call-sign-cue conditions, the 2×2 factorial design represents all combinations of type of cue (ear or call sign) and gender match between the target and competing talkers (same or different) for both selective and divided attention. When the results for the dichotic and monaural call-sign cue conditions are analyzed, then the two repeated-measures independent variables are gender match and presentation mode (monaural or dichotic) for measures of both selective and divided attention. The subject group was the only between-subject independent variable and there were two values for this variable (young normal hearing, elderly hearing impaired).

B. Results and discussion

For the control conditions, in which two blocks of 32 trials of a single talker, *without a competing talker*, were presented to each ear, all scores for each ear and each of the 13 older adults exceeded 96% and 21 of the 26 scores were 100%. Thus, with the spectral shaping used in this study, none of the EHI listeners had any difficulty identifying the correct color-number coordinates when the spectrally shaped target stimulus was presented without a competing sentence.

The individual percent-correct scores for all subjects were transformed into rationalized arcsine units (RAU; Studebaker, 1985) to stabilize the error variance. Means and standard errors for both groups of listeners and all six test conditions are shown in Fig. 3 for the measures of selective attention (top panel) and divided attention (bottom panel). As noted previously, given the unbalanced design across the six conditions in this experiment, the six conditions were divided into two overlapping sets of four conditions each, for statistical analysis. For both selective attention and divided attention, one set of four conditions included those for which

effects of cue type $[F(1,21)=67.7]$ and gender match $[F(1,21)=26.5)$ were observed such that scores were generally higher in dichotic listening for the ear cue than for the call-sign cue and for different talker genders rather than same talker genders. In addition, the main effect of the subject group was not significant $[F(1,21)=2.8]$. There was, however, a significant interaction between gender match and group $[F(1,21)=9.1]$. An inspection of the data in the left and center portions of the top panel reveals that the older subjects show a greater increase in performance than the younger subjects when gender match changes from the same to different.

In the corresponding GLM analysis involving divided attention and dichotic listening conditions (left and center portions of lower panel in Fig. 3), significant main effects of cue type $[F(1,21)=110.5]$ and subject group $[F(1,21)=20.8]$ were observed, but the effect of gender match was not significant $[F(1,21)=0.9]$. Two interactions, however, were found to be significant as well, both involving the subject group. Specifically, subject group interacted significantly with cue type $[F(1,21)=11.0]$ and gender match $[F(1,21)=5.6]$. Regarding the interaction of group and cue type, the ear cue yielded higher scores than the call-sign cue, but the difference was larger for the young, normal-hearing listeners. With regard to the interaction of group with gender match, there appears to be a slight effect of gender match for the older subjects only, such that performance was higher when the gender of the target and competing talkers differed.

In summary, for both young and older adults and selective and divided attention, the ear cue yielded higher performance in dichotic listening than the call-sign cue. In addition, for both selective and divided attention, the gender match between target and competing talkers was a significant factor in dichotic listening for the older adults such that scores were higher in these subjects when the two talkers differed in gender. It is interesting that gender match of the two talkers had a significant effect for the EHI listeners, even when the two talkers were delivered to opposite ears. This clearly cannot be an issue related to increased "energetic" masking in the EHI group, as there is no physical or peripheral interaction among the target and competing stimuli when delivered to opposite ears.

To examine group differences for each of the four dichotic conditions, a series of independent-group t tests were performed. Here, given four comparisons, the criterion p value of 0.05 was divided by 4 and $p < 0.0125$ was established as the criterion for significance (i.e., p was adjusted for multiple comparisons). Using this significance criterion, none of the group comparisons were significant for selective attention, but three of the four group comparisons were significant for divided attention. These significant group differences have been marked in the left and center panels of Fig. 3 with asterisks.

2. Effects of mode of presentation, gender match, and group

A parallel set of analyses were performed for the four conditions in the center and right portions of Fig. 3 for both selective attention (top panel) and divided attention (bottom

FIG. 3. Means and standard errors (error bars) of the transformed percent-correct scores for the CRM for each of the 12 listening conditions in experiment 1 and both groups of subjects (grey bars: YNH; black bars: EHI). Asterisks mark significant ($p<0.01$) differences between the two groups.

dichotic stimulus presentation was employed (left and center portions of each panel in Fig. 3), whereas the other set of four conditions included those for which the call sign was the cue (center and right portions of each panel in Fig. 3). A total of four mixed-model General Linear Model (GLM) analyses were conducted using a 2×2 factorial design for the repeated-measures variables and a between-subject factor of group (YNH or EHI). In the first two analyses, the repeated-measures variables were cue type (ear versus call sign) and gender match (same versus different), with separate analyses performed on the measures of selective and divided attention. In the next two analyses, the repeated-measures variables were presentation mode (monaural versus dichotic) and gender match (same versus different), with separate analyses again performed for measures of selective and divided attention.

1. Effects of cue type, gender match, and group

For the first GLM analysis involving selective attention in dichotic listening conditions, significant ($p<0.01$) main

panel). In this case, significant main effects were observed for each of the three independent variables [presentation mode: $F(1,21)=422.5$; gender match of talkers: $F(1,21)=190.4$; and group: $F(1,21)=7.6$], but one two-way and one three-way interaction involving various combinations of these variables were also significant. Specifically, the interaction between the presentation mode and gender match [$F(1,21)=75.7$] and the three-way interaction of presentation mode, gender match, and group [$F(1,21)=11.1$] were significant. Overall, from an inspection of Fig. 3, these results indicate that there was a significant effect of gender match in both presentation modes, but it was larger for monaural presentations and the younger listeners. Clearly, this complex pattern of significant findings is driven by the extremely low performance of both subject groups for the monaural presentation mode when the target and competing talkers were the same gender. Finally, the asterisk in the right portion of the top panel of Fig. 3 indicates that the older adults performed more poorly than the young adults in this condition (independent-sample t test with p values adjusted for multiple comparisons, as previously).

For divided attention (the center and right portions of the bottom panel of Fig. 3), parallel GLM analyses revealed identical statistical findings as those reported for selective attention. Specifically, significant main effects were observed for each of the three independent variables [presentation mode: $F(1,21)=106.0$; gender match of talkers: $F(1,21)=43.3$; and group: $F(1,21)=16.0$], but one two-way and one three-way interaction involving various combinations of these variables were also significant. The interaction between presentation mode and gender match [$F(1,21)=50.9$] and the three-way interaction of the presentation mode, gender match, and group [$F(1,21)=12.8$] were significant. As before, from an inspection of the bottom panel of Fig. 3, these results indicate that there was a significant effect of gender match in both presentation modes, but it was larger for monaural presentations and the younger listeners. Once again, this complex pattern of significant findings appears to be driven by the low performance of both subject groups for the monaural presentation mode when the target and competing talkers are the same gender. This clearly was the most challenging listening situation for both groups of listeners and for both the selective- and divided-attention tasks. Finally, the asterisks in the center and right portions of the bottom panel of Fig. 3 indicate that the older adults performed more poorly than the young adults for three of the four divided-attention conditions (independent-sample t tests with p values adjusted for multiple comparisons).

In Fig. 3, note that the data for the monaural, call-sign cue, same-gender listening condition revealed the lowest performance for both groups and both types of attention. In this listening condition, it is most difficult to segregate the target and competing stimuli and both groups perform poorly when it is difficult to perform this segregation. When segregation of the target and competing stimuli has been facilitated through differences in talker gender (monaural presentation, call-sign cue, different gender) or laterality/location (dichotic presentation, call-sign cue, same gender), the relative improvements in performance are more sizable for the young

adults than for the older adults. Thus, the younger adults appear to benefit more than the older adults, on average, from the segregation of the target and competing messages.

3. Correlations and individual differences

Correlational analyses of the individual data for all 23 subjects were performed for the six selective-attention measures and the six divided attention measures. Correlations were computed for each group separately as well, but the patterns of correlations were very similar for each group and for both groups combined so only the latter are summarized here. For the 6 measures of selective attention there are 15 possible pairwise correlations, and 10 of these were positive, strong ($r>0.75$ in 9 of the 10 cases), and statistically significant ($p<0.01$). All five nonsignificant correlations were between performance measured for the monaural condition using the call-sign cue and the same gender for target and competing talkers and performance measured for the other five conditions. That is, performance in this very difficult condition was not correlated with performance in any of the other conditions. Performance in the other five conditions, however, was strongly correlated such that those who did well in one of those five conditions tended to do well in the others.

For the divided-attention conditions, the pattern of correlations was somewhat similar to that observed for selective attention. The five correlations with performance for the monaural listening, call-sign cue, same-gender condition were once again lower than the rest, but here the correlations were stronger ($0.46<r<0.71$) than observed for selective attention with four of the five being statistically significant. Further, for the remaining ten correlations involving performance measured in the other five conditions, all were significant and the correlations ranged from 0.78 to 0.94. All told, 14 of the 15 correlations for divided attention were significant and 13 of them were at least 0.70.

Six correlations were also calculated between the measures of selective and divided attention; one correlation for each of the six test conditions. These correlations ranged from 0.51 to 0.83 with five of the six being significant ($p<0.01$). Thus, those who did better on the measures of selective attention tended to do better on the measures of divided attention, at least for conditions that were otherwise identical (presentation mode or cue type and gender match).

Correlations were also computed between each of the six measures of selective and divided attention and age, average high-frequency hearing loss (1000, 2000, and 4000 Hz) in both ears, hearing loss asymmetry (the average high-frequency hearing loss in the right ear minus that in the left ear), and average digit-span score (average of forward and backward digit span measures) for the 13 older adults. It should be noted that there were no significant correlations among these four potential "explanatory" variables themselves. For the six measures of selective attention, there was only one significant ($p<0.05$) correlation observed. This was a negative correlation ($r=-0.62$) between hearing loss asymmetry and performance in the ear-cue, same-gender condition. However, the correlation with hearing loss asymmetry approached significance ($r=-0.52$, $p=0.07$) for the
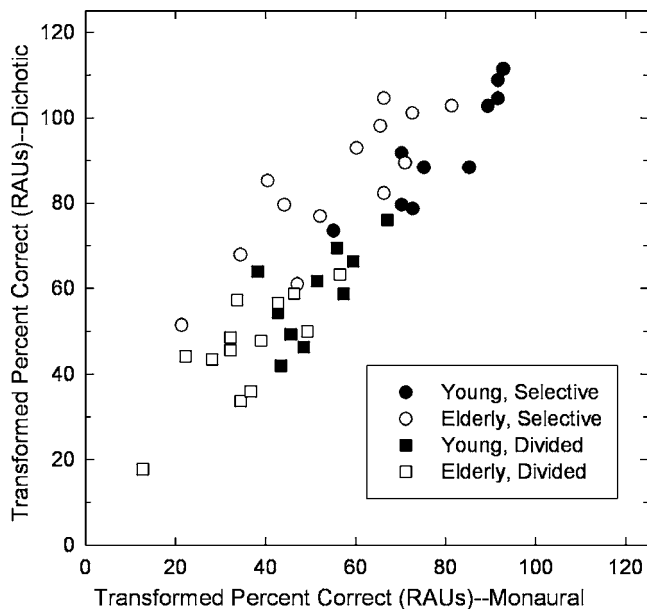
FIG. 4. Scatter plot of transformed percent-correct CRM scores for dichotic listening conditions as a function of monaural listening for the same conditions. The call sign was always the cue and only conditions making use of different talker gender are included in this figure. Data from the 10 YNH subjects are shown by filled symbols and for the 13 EHI subject by the unfilled symbols. Circles represent values for selective-attention conditions and squares for divided-attention conditions. For the data from the YNH listeners, the Pearson-r correlation coefficient between monaural and dichotic performance was 0.94, whereas it was 0.87 for the EHI listeners.

corresponding condition with the call-sign cue. In addition, two other moderate correlations ($r=0.53$ and 0.54) for measures of selective attention approached significance ($p=0.06$) and these were with the digit-span measure. For divided attention, the only significant correlations were with digit span. Four of the six measures of divided attention were significantly correlated ($0.61 < r < 0.77$) with digit span and the other two ($r=0.54$ and 0.52) approached significance ($p=0.06$ and 0.07, respectively).

The foregoing analysis of individual data suggests that there may be common mechanisms underlying the identification of speech from one talker with another talker speaking in the background, regardless of whether the competing talkers are presented to the same ear (monaural) or to different ears (dichotic). Further, these two attentional mechanisms (selective and divided) are strongly correlated.

The scatter plot in Fig. 4 further emphasizes that the differences in performance on dichotic selective- and divided-attention tasks are closely related to similar differences observed for monaural presentations of these same stimuli. In this figure, only the data for the different-gender conditions and call-sign cue are presented. The same-gender conditions were very difficult for monaural presentations, not correlated with other measures, especially for selective attention, and likely reflected an inability of the subject to segregate the two messages. As such, it is probably not appropriate to examine attention for two competing stimuli if the stimuli cannot be perceived as two separate stimuli. In addition, the scatter plot was generated only for conditions using the call-sign cue because that is the only cue in common across the monaural and dichotic presentation conditions. In

Fig. 4, note that, for both age groups, as performance in either the selective- or divided-attention task increases for monaural presentation (x axis), performance also increases for the corresponding task with dichotic presentation (y axis). In fact, the correlation coefficient between scores for monaural and dichotic conditions is 0.94 for young adults and 0.87 for older adults. This suggests, at least for these stimuli and conditions, that the performance in dichotic conditions may reflect individual differences in the same cognitive mechanisms underlying performance in monaural conditions and not necessarily central-auditory factors, such as binaural processing, that are modality specific and unique to dichotic processing.

IV. EXPERIMENT 2

A. Experiment-specific procedures

There were a total of six conditions in this experiment; three each for selective and divided attention. All presentations were monaural (right ear) and the cue was always talker gender (male or female). The sole repeated-measures independent variable was talker uncertainty for both dependent variables (measures of selective and divided attention). There were three levels of talker uncertainty examined: minimum, medium, and maximum uncertainty. For "minimum uncertainty," the same male talker and the same female talker were used for all 32 trials in a block and for all four blocks. For "medium uncertainty," the same talker of one gender was used as one talker across trials while the other talker was randomly selected on each trial from among four possible talkers of the opposite gender. There were two variations of this "medium uncertainty" case: one in which the fixed talker was male and the variable talkers were female and one in which the fixed talker was female and the variable talkers were male. Four blocks of 32 trials were completed for each of these variations of the "medium uncertainty" conditions and the results for all eight blocks were averaged. Finally, for "maximum uncertainty" one male and one female were chosen randomly for each trial from the four possible talkers for each gender. The randomization was constrained such that each of the eight talkers appeared four times as the target talker and four times as the competing talker in a block of 32 trials. Thus, there was one repeated-measures independent variable (talker uncertainty) with three values. The subject group was again the only between-subject independent variable and there were two values for this variable (young normal hearing, elderly hearing impaired).

B. Results and discussion

1. Effects of stimulus uncertainty and group

In this experiment, the use of talker gender as a cue for selective and divided attention was examined in the same 13 elderly adults and 10 young adults, 7 of whom had participated in experiment 1. Given the use of talker gender as the cue, the genders of the target and competing talkers were always different and gender match was no longer a viable independent variable. Instead, the variation in talker voices from trial to trial, referred to here as talker uncertainty, was manipulated systematically.

Talker-Gender Cue, Monaural
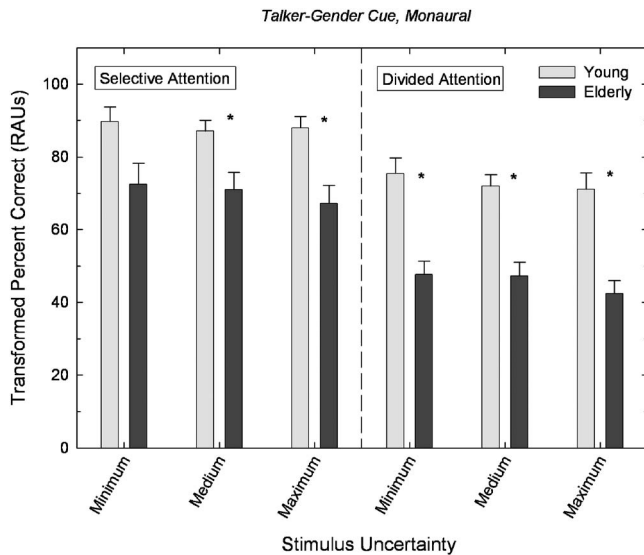
FIG. 5. Means and standard errors (error bars) of the transformed percent-correct scores for the CRM for each of the six listening conditions in experiment 2 and both groups of subjects (grey bars: YNH; black bars: EHI). Asterisks mark significant ($p < 0.01$) differences between the two groups.

Figure 5 presents the means and standard errors, in RAUs, for the three levels of stimulus uncertainty included in experiment 2, with the selective-attention measures in the left half of the figure and the divided-attention measures in the right half. Casual visual inspection of these data suggests that, in general, performance was lower in the divided-attention tasks than in the comparable selective-attention tasks and that older adults generally performed worse than young adults. Two 2×3 mixed-model GLM analyses were performed, one for the measures of selective attention and one for the measures of divided attention.

The GLM analysis on the measures of selective attention (the left half of Fig. 5) revealed a significant main effect of subject group [$F(1,21)=8.1$], but not uncertainty [$F(2,42)=2.1$]. In addition, there was no interaction between uncertainty and group [$F(2,42)=1.0$]. Multiple independent-sample t tests were computed to examine the differences in performance between groups for each uncertainty condition. The criterion p value was again adjusted for multiple comparisons ($p=0.05/3=0.0167$) and the asterisks in the left half of Fig. 5 mark those conditions for which a significant difference between groups was observed. The older adults performed significantly worse than the younger adults on two of the three selective-attention tasks, the two having greater amounts of stimulus uncertainty.

The GLM analysis on the measures of divided attention (the right half of Fig. 5) revealed a significant main effect of subject group [$F(1,21)=27.3$] and uncertainty [$F(2,42)=4.8$]. The group x uncertainty interaction was not significant [$F(2,42)=0.9$]. With regard to stimulus uncertainty, post hoc paired-sample t tests (adjusted for multiple comparisons) indicated that the mean performance for the maximum uncertainty condition was significantly lower than that for the minimum uncertainty condition. No other paired comparisons involving stimulus uncertainty were significant. To examine the effect of subject group in more detail, mul-

tiple independent-sample t tests were computed for each uncertainty condition, with the criterion p value adjusted for multiple comparisons. The asterisks in the right half of Fig. 5 mark those conditions for which a significant difference between groups was observed. The older adults performed significantly worse than the younger adults on all three divided-attention tasks.

To summarize the group data, the EHI adults performed significantly worse than YNH adults in five of the six conditions included in this experiment. These findings are consistent with the findings of experiment 1, but using talker gender as the cue rather than the call sign or ear. The results from this experiment with regard to uncertainty were less clear. There was no effect of uncertainty on measures of selective attention, but there was on measures of divided attention. Although a statistically significant effect of stimulus uncertainty was observed between the minimum and maximum uncertainty conditions for the measures of divided attention, the mean differences in performance over this range of uncertainty were not large (4-5 RAU) for either group. This small effect could be due to the restricted range of talker uncertainty available for the CRM materials used in this experiment. Moreover, a significant interaction between the effects of uncertainty and group was not observed for either measure of attention suggesting that the effects of uncertainty, or lack thereof, were equivalent for both young normal-hearing subjects and elderly hearing-impaired listeners. As noted previously, others have found older adults to be more vulnerable to such talker uncertainty or variability in other contexts (Sommers, 1997), although a much wider range of uncertainty was encompassed in that study than in this experiment. In addition, Brungart and Simpson (2004) found no significant effects of talker uncertainty on the CRM task in young adults, whereas Coughlin (2004) observed significant effects of talker uncertainty in both young and older adults on the CRM task (for selective attention).

2. Correlations and individual differences

With regard to the individual data from experiment 2, correlations across the three stimulus uncertainty conditions for each measure of attention were all strong, positive, and statistically significant ($p < 0.05$). Specifically, the three pairwise correlations for selective attention were 0.91, 0.89, and 0.92 and those for divided attention were 0.95, 0.91, and 0.93 for the entire set of 23 subjects. Similar correlations were also observed within each group separately. Finally, when the three measures of selective attention were correlated with the three measures of divided attention for all 23 subjects, the nine correlations ranged from 0.68 to 0.87 and were all statistically significant. Thus, an individual who performed well on any of the tasks in this experiment tended to perform well on all of the tasks, regardless of the type of attention being measured or the amount of stimulus uncertainty.

Correlations were also computed between each of the six measures of selective and divided attention and age, average high-frequency hearing loss (1000, 2000, and 4000 Hz) in both ears, and average digit-span score (average of forward and backward digit span measures) for the 13

Humes, Lee, and Coughlin: Auditory selective and divided attention

older adults. It should be noted that there were no significant correlations among these three potential "explanatory" variables. For divided attention measures, the only significant correlation that was observed was a negative correlation ($r = -0.62$) between age and performance on the maximum uncertainty condition. Thus, among the 13 older adults, the older the subject, the lower the score for divided attention under maximum uncertainty conditions. For selective attention, all three measures of selective attention were correlated significantly with digit span ($0.57 < r < 0.64$) and there were no significant correlations with age or hearing loss among the 13 older adults. Thus, for the measures of selective attention used here, those with better short-term memory performed better than those with poorer short-term memory.

V. GENERAL DISCUSSION

Across both experiments 1 and 2, several consistent trends were observed when using the CRM and two competing talkers to measure attention in young and older adults. First, in general, for both young and older adults, divided-attention tasks yielded lower scores than comparable selective-attention tasks. This was not unexpected since divided attention is known to place greater demands on memory (Pashler, 1998). This was true for both monaural and dichotic presentation modes and all three types of attention cue used (ear, call sign, talker gender). The lone exception probably being when monaural, same-gender conditions were used in experiment 1. This was also the condition yielding overall performance for both age groups that was well below that of any other test condition in either experiment. It is likely that, in this condition, the two competing talkers were difficult to segregate given similar fundamental frequencies (Assmann and Summerfield, 1990; Culling and Darwin, 1993; Arehart, King, and McLean-Mudgett, 1997; Summers and Leek, 1998) and energetic forms of masking predominated, making it extremely difficult to hear out or segregate the two auditory stimuli and respond accordingly.

A second general trend across experiments was related to the difference between the performance of young and old adults. In particular, in eight of the nine divided-attention measures across both experiments, the EHI subjects performed significantly worse than the YNH subjects. This was true regardless of the presentation mode, the type of cue used, the gender match between target and competing talkers, or the amount of talker uncertainty. By comparison, significant differences in performance between age groups were only observed on three of the nine selective-attention tasks.

Another common finding across both experiments pertained to the strong correlation among CRM scores for a given experiment, suggesting considerable redundancy and a common underlying factor or factors. There were consistent individual differences in performance in both experiments such that high performers tended to score high on all conditions and lower performers tended to score low on all conditions. Subsequent correlations between performance and age, digit span, and hearing loss for the 13 older adults suggested that, at least for this age group, the individual differences in performance were associated frequently with indi-

vidual differences in memory (digit span). As noted in the Introduction, the observations of age differences in performance on selective- and divided-attention tasks, as well as the association of individual differences in performance with memory among the older adults, have been noted before (Verhaeghen and De Meersman, 1998a, 1998b; Verhaeghen *et al.*, 2003). In our particular implementation of attention measurements in experiments 1 and 2, the subjects were required to hold the orthographic cue (ear, call sign, or gender) in memory for about 1.5–2.0 s before responding with the color-number coordinates for measures of selective attention. For divided-attention measures, the subjects had to hold the stimuli in memory for about 3 s before the orthographic cueing was completed and response collection enabled. Thus, in this study, as in previous studies, age-related differences in attention could be mediated by corresponding differences in memory.

The absence of other correlations, with age, hearing loss, or hearing loss asymmetry (with one exception in experiment 1), is perhaps as important as these significant correlations. Among the elderly, differences in age or differences in hearing loss did not contribute to individual differences in performance from either experiment. Use of spectral shaping to minimize audibility limitations appeared to be effective, as was also the case in Coughlin (2004), although a slightly different approach was taken to ensure audibility through 4000 Hz in the previous study.

The percentage of errors in each experiment, for each type of attention and each group of subjects, that resulted from intrusions of the color or number coordinates, or both, from the competing stimulus was examined. An "intrusion" of the competing color-number coordinates simply means that the listener's response included the color, number, or color-and-number coordinates from the competing talker. GLM analyses of the percentage of intrusions for each experiment, revealed only significant ($p < 0.01$) main effects of type of attention in each experiment, but no significant ($p > 0.05$) differences between groups or interactions between group and type of attention. Selective attention revealed a greater percentage of intrusions than divided attention. Since the competing response coordinates could not have the same color or number as the target, on any given trial, that left three possible colors and seven possible numbers for use as coordinates in the competing message. Thus, there is a 1/3 chance the color alone from the competing stimulus would be reported by guessing, a 1/7 chance for the number alone, and a 1/21 chance that both the color and number of the competing stimulus would be reported randomly. Summing all these possible intrusion probabilities results in a total intrusion percentage of errors of 11/21 (i.e., (1/3+1/7 +1/21) or 52.4%.

For color-only intrusions, the percentage of errors that were of this type were not unlike those expected by chance, except for the divided attention conditions for the EHI subjects. In this case, the percentage of color-only intrusions was quite a bit less than would be expected by random guessing. For number-only intrusions, the percentages of errors that were of this type were fairly consistent across all subject and attention combinations and were also not unlike
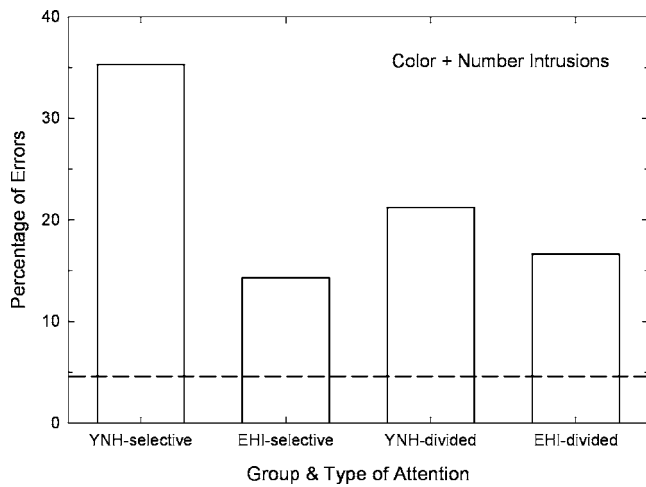
FIG. 6. The mean percentage of errors due to color+number intrusions. The percentage of intrusions expected by chance is represented by a horizontal dashed line.

those expected by chance. The situation was considerably different, however, for the color+number intrusions and these intrusions are illustrated in Fig. 6. As shown, the percentages of errors that were of this type were about three to five times greater than those expected by chance. The analysis of the percentage of errors that are intrusions from the competing stimulus, especially for the color+number intrusions, provides further support for the notion that the underlying processing reflects attentional capabilities. The listeners were less able to tune out, inhibit, or suppress the competing coordinates and these coordinates surface as responses in the errors much more frequently than would be expected by chance alone. It has frequently been shown that older adults are less able to inhibit irrelevant or competing sensory information (e.g., Hasher and Zacks, 1988; Hasher *et al.*, 2006), including the situation of attending to a target talker and ignoring a competing talker (Sommers and Danielson, 1999; Tun, O'Kane, and Wingfield, 2002), although age differences in the interference of a competing talker with the target talker have not always been observed (e.g., Li *et al.*, 2004). In the present study, however, although the older adults made more errors than the young adults in most conditions, the relative percentage of color + number intrusion errors was lower for older adults (Fig. 6). Thus, it was more often the case that *neither* the target or competing coordinates were identified correctly by the older adults whereas younger adults were superior at recalling both (incorrectly in the case of color + number intrusions).

It should be kept in mind that the target-to-competition ratio in this study was 0 dB and, for the EHI listeners, both stimuli underwent the same amount of spectral shaping. Overall, therefore, the target and competing stimuli are equivalent in energetic terms and it is impossible to state which stimulus is the energetic "masker" and which is the "signal." That is, it is not the case that the competing coordinates emerged more frequently among the errors because the competing stimulus masked, in an energetic sense, the target coordinates. Rather, given acoustically, and presum-

ably perceptually, equivalent target and competing stimuli, attention and memory processes determined both the percentage of target coordinates recalled correctly *and* the large percentage of the errors attributable to the competing coordinates.

The use of the same stimuli and same call-sign cue for four dichotic and four monaural conditions in experiment 1 enabled a comparison of the individual differences in performance for both presentation modes (Fig. 4). Strong correlations of about 0.9 between monaural and dichotic performance for selective and divided attention and in young and elderly adults strongly suggests that there are common underlying mechanisms for these two presentation modes. It is suggested here that the common underlying mechanism is attention or attention+memory; in either case, a cognitive process. Thus, individual differences in performance on other dichotic speech-identification or speech-recognition tasks may also be attributable to underlying individual differences in attention. As such, these measures of dichotic speech processing should not be considered to be pure measures of "central auditory" processing, especially when used with older adults (e.g., Jerger *et al.*, 1989; Hallgren *et al.*, 2001; Humes, 2005).

Finally, the mean CRM scores for selective and divided attention conditions across experiments 1 and 2 for both age groups (Figs. 3 and 5) were compared. Mean values were computed for each group and for the four cue/presentation mode conditions used in the two experiments. For experiment 1, only the conditions using talkers of different genders were averaged since different genders were always used in experiment 2. In addition, since the trial-to-trial uncertainty in experiment 1 was most like the medium-uncertainty conditions from experiment 2, the medium-uncertainty condition from that experiment was used in this across-experiment comparison. From these across-experiment comparisons, the spatial/laterality (ear) cue in dichotic listening yielded the highest mean performance for both age groups and for both selective and divided attention. Likewise, the call-sign cue in the monaural listening condition yielded the lowest performance within each attention type and for both groups. The ordering of the two middle cues/conditions, however, was not as clear. For selective attention, for example, the call-sign cue in dichotic listening tended to produce somewhat higher scores than the gender cue in monaural listening conditions, whereas there was no such trend, or even an opposite trend, for these two cue/condition combinations in the divided-attention task. Caution should be exercised when comparing performance across cues and conditions, however, because, despite attempts to equate various factors across experiments and conditions, this has not been fully accomplished. For example, for the call sign as a cue, there were eight alternatives, whereas for the other two cues, gender and ear, there were only two alternatives. In that sense, there was more uncertainty with regard to the potential cues from trial to trial when call sign was used. To examine the role of cue uncertainty, additional experiments with call sign restricted to only two choices, just as gender and ear, would be required. Nonetheless, it appears that those cues associated with physical aspects of the perceived auditory object,

Humes, Lee, and Coughlin: Auditory selective and divided attention

such as its location (ear) or fundamental frequency (gender), are stronger than the arbitrary lexical labels (call signs) attached to those objects.

American National Standards Institute. (1996). "Specifications for audiometers," ANSI S3.6-1996, American National Standards Institute, New York, NY.

American National Standards Institute. (1999). "Maximum permissible ambient noise levels for audiometric test rooms," ANSI S3.1-1999, American National Standards Institute, New York, NY.

Arehart, K. H., King, C. A., and McLean-Mudgett, K. S. (1997). "Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss," J. Speech Lang. Hear. Res. 40, 1434–1444.

Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. 88, 680–697.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. 107, 1065–1066.

Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. 109, 1101–1109.

Brungart, D. S., and Simpson, B. D. (2004). "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," J. Acoust. Soc. Am. 115, 301–310.

CHABA (1988). "Speech understanding and aging," J. Acoust. Soc. Am. 83, 859–895.

Coughlin, M. P. (2004). "Aided speech recognition in single-talker competition by elderly hearing-impaired listeners," Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana.

Cox, R. M., and McDaniel, D. M. (1986). "Reference equivalent threshold levels for pure tones and 1/3-octave noise bands: Insert earphone and TDH-49 earphone," J. Acoust. Soc. Am. 79, 443–446.

Culling, J. F., and Darwin, C. J. (1993). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by fundamental frequency," J. Acoust. Soc. Am. 93, 3454–3467.

Divenyi, P. L., and Haupt, K. M. (1997a). "Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects," Ear Hear. 18, 42–61.

Divenyi, P. L., and Haupt, K. M. (1997b). "Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. II. Correlational analysis," Ear Hear. 18, 100–113.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician," J. Psychiatr. Res. 12, 189–198.

Hallgren, M., Larsby, B., Lyxell, B., and Arlinger, S. (2001). "Cognitive effects in dichotic speech testing in elderly persons," Ear Hear. 22, 120–129.

Hasher, L., and Zacks, R. (1988). "Working memory, comprehension, and aging: A review and a new view," in The Psychology of Learning and Motivation, edited by G. H. Bower (Academic Press, New York), pp. 129–225.

Hasher, L., Lustig, C., and Zacks, R. (2006). "Inhibitory mechanisms and the control of attention," in edited by A. Conway, C. Jarrold, M. Kane, A. Miyake, and J. Towse, Variation in Working Memory (Oxford University Press, London) (in press).

Helfer, K. A., and Wilber, L. A. (1990). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech Hear. Res. 33, 149–155.

Humes, L. E. (1991). "Prescribing gain characteristics of linear hearing aids," in The Second Vanderbilt/VA Hearing Aid Report, edited by G. A. Studebaker, F. H. Bess, and L. Beck (York Press, Parkton, MD).

Humes, L. E. (1996). "Speech understanding in the elderly," J. Am. Acad. Audiol 7, 161–167.

Humes, L. E. (2002). "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," J. Acoust. Soc. Am. 112, 1112–1132.

Humes, L. E. (2005). "Do "auditory processing" tests measure auditory processing in the elderly?," Ear Hear. 26, 109–119.

Humes, L. E., and Roberts, L. (1990). "Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility," J. Speech Hear. Res. 33, 726–735.

Humes, L. E., and Christopherson, L. (1991). "Speech-identification difficulties of hearing-impaired elderly persons: the contributions of auditory-processing deficits," J. Speech Hear. Res. 34, 686–693.

Humes, L. E., and Halling, D. C. (1993). "Overview, rationale and comparison of suprathreshold-based gain-prescription procedures," in Strategies for Selecting and Verifying Hearing Aid Fittings, edited by M. Valente (Thieme Medical, New York).

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (1994). "Factors associated with individual differences in clinical measures of speech recognition among the elderly," J. Speech Hear. Res. 37, 465–474.

Jerger, J., Jerger, S., Oliver, T., and Pirozzolo, F. (1989). "Speech understanding in the elderly," Ear Hear. 10, 79–89.

Jerger, J., Jerger, S., and Pirozzolo, F. (1991). Correlational analysis of speech audiometric scores, hearing loss, age and cognitive abilities in the elderly, Ear Hear. 12, 103–109.

Lee, L. W., and Humes, L. E. (1993). "Evaluating a speech-reception threshold model for hearing-impaired listeners," J. Acoust. Soc. Am. 93, 2879–2885.

Li, L., Daneman, M., Qi, J., and Schneider, B. A. (2004). "Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. 30, 1077–1091.

McDowd, J. M., and Shaw, R. J. (2000). "Attention and aging: A functional perspective," In The Handbook of Aging and Cognition, 2nd ed., edited by F. I. M. Craik and T. A. Salthouse (Erlbaum, Mahwah, NJ) pp. 221–292.

Pashler, H. E. (1998), The Psychology of Attention (MIT Press, Cambridge, MA).

Rogers, W. A. (2000). "Attention and aging," in Cognitive Aging: A Primer, edited by D. C. Park and N. Schwarz (Psychology Press, Philadelphia, PA) pp. 57–73.

Sommers, M. S. (1997). "Stimulus variability and spoken word recognition. II: The effects of age and hearing impairment," J. Acoust. Soc. Am. 101, 2278–2288.

Sommers, M. S., and Danielson, S. M. (1999). "Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context," Psychol. Aging 14, 458–472.

Souza, P. E., and Turner, C. W. (1994). "Masking of speech in young and elderly listeners with hearing loss," J. Speech Hear. Res. 37, 655–661.

Studebaker, G. A. (1985). "A rationalized arcsine transform," J. Speech Hear. Res. 28, 455–462.

Summers, V., and Leek, M. R. (1998). "F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," J. Speech Lang. Hear. Res. 41, 1294–1306.

Tun, P. A., O'Kane, G., and Wingfield, A. (2002). "Distraction by competing speech in younger and older listeners," Psychol. Aging 17, 453–467.

van Rooij, J. C. G. M., and Plomp, R. (1990). "Auditive and cognitive factors in speech perception by elderly listeners. II. Multivariate analyses," J. Acoust. Soc. Am. 88, 2611–2624.

van Rooij, J. C. G. M., and Plomp, R. (1992). "Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion," J. Acoust. Soc. Am. 91, 1028–1033.

van Rooij, J. C. G. M., Plomp, R., and Orlebeke, J. F. (1989). "Auditive and cognitive factors in speech perception by elderly listeners. I. Development of test battery," J. Acoust. Soc. Am. 86, 1294–1309.

Verhaeghen, P., and De Meersman, L. (1998a), "Aging and the negative priming effect: A meta-analysis," Psychol. Aging 13, 1–9.

Verhaeghen, P., and De Meersman, L. (1998b). "Aging and the Stroop effect: A meta-analysis," Psychol. Aging 13, 120–126.

Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., and Cerella, J. (2003). "Aging and dual-task performance: A meta-analysis," Psychol. Aging 18, 443–460.

Walden, T. C., and Walden, B. E. (2004). "Predicting success with hearing aids in everyday living," J. Am. Acad. Audiol 15, 342–352.

New speech harmonic structure measure and its applications to speech processing

An-Tze Yu[a)]

Department of Computer Science, National Chupei Senior High School, Hsinchu, Taiwan, ROC,
No. 3, Jungyang Rd., Jubei City, Hsinchu, Taiwan 302, ROC

Hsiao-Chuan Wang[b)]

Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, ROC,
101, Sec. 2, Kuang Fu Road, Hsinchu, Taiwan, 30055, ROC

The harmonic structure can be easily recognized in time-frequency representation of speech signals in adverse environment. The harmonicity is a measure of the completeness of a harmonic structure. This paper presents a new harmonic structure measure that extends the conventional harmonicity to a set of harmonicities. They are expressed in terms of the grid harmonicity, the temporal harmonicity, the segment-spectral harmonicity, and the segmental harmonicity. The grid harmonicity measures the completeness of individual harmonics in each frame. The grid harmonicities in a frame are summed up to form a temporal harmonicity for representing the strength of harmonicity. The segment-spectral harmonicity, computed by summing specific grid harmonicity over a segment, evaluates the integrity of individual harmonics across a segment. The segmental harmonicity evaluates the total strength of harmonic structure within a segment. This set of harmonicities is available for a systematic analysis of the harmonic structure and effective to several speech processing tasks. The applications to speech distortion analysis, robust fundamental frequency estimation, robust voicing detection, and speech enhancement are demonstrated. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2356838]

PACS number(s): 43.72.Ar, 43.72.Ne [DOS]   Pages: 2938–2949

I. INTRODUCTION

The harmonic structure can be easily recognized in time-frequency representation of speech signals even in environments of very low signal-to-noise ratio (SNR). The harmonicity denotes the degree to which a signal is harmonic or periodic. For speech signals, harmonicity is synonymous with the degree of voicing. Many methods have been derived to evaluate the harmonic structure of speech signals. The normalized autocorrelation is a simple yet effective method to estimate the periodicity of a signal. The harmonics-to-noise ratio (HNR) measured in the autocorrelation domain is a useful method for estimating the harmonicity (Boersma, 1993). Some cepstrum-based techniques are used to provide more efficient algorithm for determining the HNR (Krom, 1993; Qi and Hillman, 1997; Murphy, 2000). The HNR can also be implemented in the instantaneous frequency amplitude spectrum (IFAS) (Arifianto and Kobayashi, 2003).

Though these methods are good enough for HNR determination and fundamental frequency (F0) estimation, they do not provide sufficient information for further analysis of the harmonic structure, for example, investigating the distortion of individual harmonics within a frame or a segment. This paper presents a work that extends conventional harmonicity measures to a set of harmonicities. They are expressed in terms of the grid harmonicity, the temporal harmonicity,

the segment-spectral harmonicity, and the segmental harmonicity. The grid harmonicity is the lower-level harmonicity that measures the completeness of individual harmonics in each frame. Other harmonicities are derived from the grid harmonicity at selected time, frequency, and segmental indices. This set of harmonicities permits a systematic analysis of the harmonic structure. They can be applied to numerous speech processing tasks, such as the speech distortion analysis, the robust fundamental frequency (F0) estimation, the robust voicing detection, and the speech enhancement.

When the speech is corrupted by an additive noise, its harmonic structure is partially distorted. The type and level of corrupting noise determine the degree and distribution of the distortion. The proposed harmonicities represent the completeness of the harmonic structure of a speech signal in different domains. Notably, the lower harmonicity corresponds to larger distortion or worse speech quality. These harmonicities provide the accurate and detailed information in the speech distortion analysis.

The fundamental frequency (F0) estimation is necessary for many applications, such as in speech coding, speech recognition, speech enhancement, and prosody extraction. Numerous F0 estimation algorithms have been developed (de Cheveigné and Kawahara, 2001; Kasi and Zahorian, 2002). However, many of those algorithms suffer from noise corruptions. Nevertheless, the validation of F0 candidates using harmonic measure can easily eliminate those incorrect F0 candidates and thereby increase the robustness of F0 estimation. This comes from the fact that the harmonic structure

[a)]Electronic mail: yuat@ms25.hinet.net
[b)]Electronic mail: hcwang@ee.nthu.edu.tw

associated with the correct F0 maximizes the harmonicity. The robust property of harmonic structure can be used for voicing detection. It provides a tool for extracting the voiced segments in noisy speech processing.

The knowledge of the harmonic structure of voiced speech should be integrated into an enhancement system to further improve the speech quality. The comb filtering and the sinusoidal modeling are well-known approaches for high-quality speech reproduction (Nehorai and Porat, 1986; McAulay and Quatieri, 1986). Rather than designing a new enhancement algorithm, a two-stage scheme, including a front enhancement process and a postenhancement process, is proposed for the speech enhancement. Any existing enhancement algorithm can be a front enhancement process. The detailed speech information obtained from harmonic analysis is then used in the postenhancement process to further improve the speech quality. The postenhancement process is designed to further reduce the noise and smooth the discontinuity of speech harmonics. The comb filter with its gain and bandwidth adapting to factors derived from temporal and segment-spectral harmonicities provides an example of this application.

The remainder of this paper is organized as follows. In Sec. II, new harmonic structure measures are introduced. The applications of proposed harmonicities to the speech distortion analysis, the robust F0 estimation, the voicing detection, and the speech enhancement are respectively described in Secs. III–VI. Some experimental results are demonstrated. Section VII gives the discussion and conclusion.

II. NEW HARMONIC STRUCTURE MEASURES

To provide a systematic harmonic structure analysis, a set of harmonicities is proposed. They include grid harmonicity, temporal harmonicity, segment-spectral harmonicity, and segmental harmonicity. The grid harmonicity measures the completeness of individual harmonics at every frame. Grid harmonicities in a frame are summed up to form a temporal harmonicity for representing the strength of harmonicity. The segment-spectral harmonicity evaluates the integrity of individual harmonics across a segment. The segmental harmonicity measures the whole strength of harmonic structure within a segment. The formulations of these harmonicities are described as follows.

A. Grid harmonicity

The grid harmonicity measures the magnitude ratio of the harmonics to its surrounding noise. It represents the completeness of an individual harmonics in a frame. The formulation of the grid harmonicity involves three components: (1) the local spectral dominance, (2) the temporal variation factor, and (3) the neighbor dominance. These three components confirm the existence of a spectral maximum in spectral domain for a voiced speech frame.

1. Local spectral dominance

Let $e(n,m)$ be the sum of weighted magnitude spectral densities around the mth harmonics at the frequency bin index mk_0. The summation range is $[(m-\eta)k_0, (m+\eta)k_0]$. Then $e(n,m)$ is expressed as follows,

$$e(n,m) = \sum_{k=\lfloor(m-\eta)k_0\rfloor}^{\lceil(m+\eta)k_0\rceil} S(n,k)\phi(k-mk_0). \tag{1}$$

where $S(n,k)$ denotes the magnitude spectral density of an analyzed signal in the nth frame, and k is the frequency bin index. k_0 is the frequency bin index of fundamental frequency. η is a fractional factor to specify the summation range. In this study, we set η to 0.5. $\phi(k)$ is a harmonics selecting function defined by the following equation,

$$\phi(k) = e^{-k^2/2\sigma^2}, \tag{2}$$

where σ^2 denotes a factor controlling the width of the harmonics selecting function. σ is chosen to be proportional to the number of frequency bins in the summation range. In this study σ^2 is set to 100. Then the local spectral dominance of mth harmonics evaluated in the nth frame is defined by

$$h_D(n,m) = \frac{e(n,m)}{\sum_{k=\lfloor(m-\eta)k_0\rfloor}^{\lceil(m+\eta)k_0\rceil} S(n,k)(1-\phi(k-mk_0))}. \tag{3}$$

Equation (3) shows that $h_D(n,m)$ is large if the harmonic does exist at mk_0. Contrarily, if the signal is nonharmonic, $h_D(n,m)$ will be a small value. Therefore, we consider $h_D(n,m)$ as a factor to evaluate the existence of a harmonic. If the speech signal is sampled at 16 kHz and transformed to the frequency domain by using fast Fourier transform (FTT) with 1024 points, the resolution in frequency domain is 15.626 Hz per frequency bin. The fundamental frequency of the speech signal is in the range of 80 to 400 Hz so that the range of k_0 is about from 5 to 25. The summation range, $[(m-\eta)k_0, (m+\eta)k_0]$, should contain at least 5 points when η is set to 0.5. In this summation range, $1-\phi(k-mk_0)$ is nonzero except at the point of $k=mk_0$. The denominator of Eq. (3) is a sum of at least four nonzero terms. Since these nonzero terms are nonnegative, the situation of singularity cannot happen.

2. Temporal variation factor

Since the harmonic structure is stable during a short time period, we can expect a small change of spectral magnitude between consecutive frames. This property can be measured by the temporal variation defined as follows,

$$D_T(n,m) = \frac{|S(n,mk_0) - S(n-1,mk_0)|}{S(n,mk_0) + S(n-1,mk_0)}$$
$$+ \frac{|S(n+1,mk_0) - S(n,mk_0)|}{S(n+1,mk_0) + S(n,mk_0)}. \tag{4}$$

This temporal variation can be converged to a value ranging between 0 and 1 by using a sigmoid function. Then a temporal variation factor is formulated as

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

A. Yu and H. Wang: New speech harmonic structure measure   2939
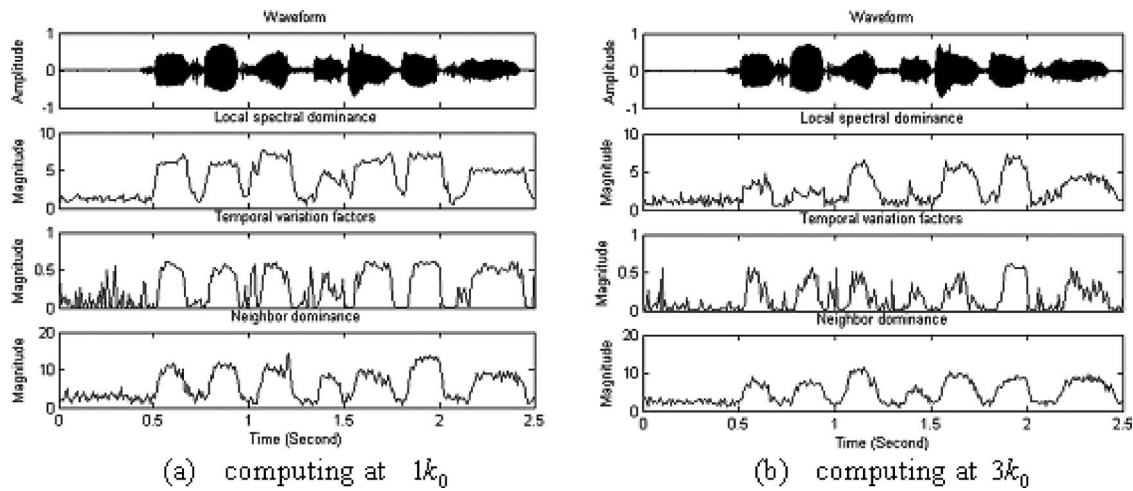
FIG. 1. Correlations of three components in computing Grid harmonicity: The sentence is spoken in Mandarin Chinese, "qi2 shi2 ta1 xi3 huan1 he1 cha2 (In fact she likes tea)."

$$h_T(n,m) = \frac{1}{1 + e^{a(D_T(n,m)-b)}}, \qquad (5)$$

where a and b define the property of transformation. In this study a and b are empirically set to 5 and 0.3, respectively. Accordingly, we can see that the temporal variation factor is large when the variation of spectral magnitude between consecutive frames is small. This implies a large temporal variation factor corresponds to a high possibility of the presence of harmonics.

3. Neighbor dominance

A number of harmonics should appear in the spectrum for a harmonic signal, and the nearby harmonics should be predictable. For this reason, we define another factor by the following equation to detect the existence of a harmonic:

$$h_N(n,m) = h_D(n,m-1) + h_D(n,m+1). \qquad (6)$$

The spectral dominance at mth harmonics should be proportional to the spectral dominances of neighboring harmonics. If the harmonic exists, $h_N(n,m)$ should be a large value. Otherwise, $h_N(n,m)$ is small.

The correlation among these three components, i.e., local spectral dominance, temporal variation factor, and neighbor dominance, is demonstrated in Fig. 1. The given utterance is a sentence spoken in Mandarin Chinese which is marked by Pinyin, "qi2 shi2 ta1 xi3 huan1 he1 cha2." This sentence means "In fact, she likes tea." There are seven syllables in this sentence. We can observe clearly the seven voiced segments.

From Fig. 1, it is obvious that these three components are large in voiced regions and small in noise or unvoiced frames. The harmonics of a signal can be confirmed when three components are large. Based on the above observation, we empirically define the grid harmonicity as the product of the local spectral dominance, the temporal variation factor, and the neighbor dominance:

$$H_G(n,m) = h_D(n,m)h_T(n,m)h_N(n,m). \qquad (7)$$

This function provides an effective discrimination between harmonic and nonharmonic signals. For the speech signal, this measure can be used to discriminate the voiced and unvoiced speech. Figure 2 illustrates the spectrogram and the grid harmonicity of the given utterance in Fig. 1. It is worthwhile to note that the computation of $H_G(n,m)$ needs an available fundamental frequency k_0. A conventional F0 algorithm can be applied to estimate the initial k_0. Then a search around the initial estimate is performed to find the accurate k_0. The criterion is to maximize the sum of all $H_G(n,m)$ in the frame.

B. Temporal harmonicity

Temporal harmonicity represents the strength of harmonicity in a frame. This measure is defined by the weighted sum of grid harmonicities in a frame:
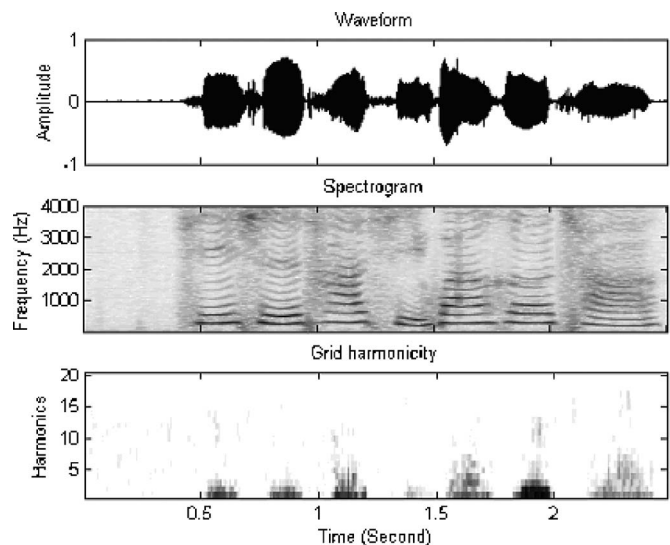
FIG. 2. Grid harmonicity of an utterance: The sentence is the same as that in Fig. 1.
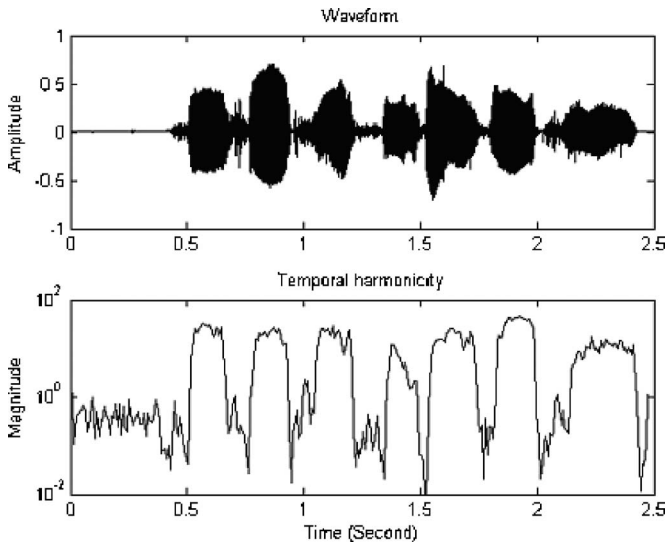
FIG. 3. Temporal harmonicity of an utterance: The sentence is the same as that in Fig. 1.

FIG. 4. Segment-spectral harmonicities of the first four segments in an utterance: segment 1—"qi2", segment 2—"shi2", segment 3—"ta1", and segment 4—"xi3".

$$H_T(n) = \sum_{m=1}^{M(n)} \alpha(n,m)H_G(n,m), \quad \alpha(n,m) = \frac{e(n,m)}{\sum_{p=1}^{M(n)} e(n,p)},$$

(8)

where $M(n)$ represents the number of harmonics in the nth frame. This integer, $M(n)$, varies with frames according to their harmonics property. In a voiced speech frame, $M(n)$ tends to be larger when the F0 is lower. $\alpha(n,m)$ plays the role of a weighting factor to $H_G(n,m)$. $H_T(n)$ is a large value if the harmonics do exist in the nth frame. It is synonymous with the degree of voicing. Figure 3 illustrates the temporal harmonicity of the same utterance as shown in Fig. 1. It shows that the temporal harmonicity is large when the frame is a voiced signal. This term provides a good measure for discriminating those voiced speech frames.

C. Segment-spectral harmonicity

The segment-spectral harmonicity measures the integrity of a specific harmonic in a speech segment. For mth harmonics existing in a segment of consecutive frames, we define a measure named segment-spectral harmonicity as the weighted sum of grid harmonicities over the segment,

$$H_S(l,m) = \sum_{n=\text{SegBegin}(l)}^{\text{SegEnd}(l)} \beta(n,m,l)H_G(n,m),$$

(9)

$$\beta(n,m,l) = \frac{e(n,m)}{\sum_{q=\text{SegBegin}(l)}^{\text{SegEnd}(l)} e(q,m)},$$

where $\text{SegBegin}(l)$ and $\text{SegEnd}(l)$ represent the first and last frame indexes belonging to the lth segment, respectively. $H_G(n,m)$ is weighted by $\beta(n,m,l)$ in computing $H_S(l,m)$. The location of a voiced segment is detected by using the temporal harmonicity. Then $\text{SegBegin}(l)$ and $\text{SegEnd}(l)$ are

determined. Figure 4 shows the segment-spectral harmonicities for first four segments of the utterance in Fig. 1. In this utterance, a segment corresponds to a syllable. We can see that the segment-spectral harmonicity is large at lower harmonics and becomes small at higher harmonics. It reveals that the lower harmonics get more complete than the higher harmonics.

The definition of segment-spectral harmonicity shown in Eq. (9) indicates that the grid harmonicity of a specific harmonic is weighted and summed over the segment. The magnitude of segment-spectral harmonicity represents the average distribution of frequency components of a segment.

D. Segmental harmonicity

Segmental harmonicity is a measure of the whole harmonic structure within a segment. This is evaluated by summing all the weighted grid harmonicities over a segment:

$$H(l) = \sum_{n=\text{SegBegin}(l)}^{\text{SegEnd}(l)} \sum_{m=1}^{M(n)} \gamma(n,m,l)H_G(n,m),$$

(10)

$$\gamma(n,m,l) = \frac{e(n,m)}{\sum_{q=\text{SegBegin}(l)}^{\text{SegEnd}(l)} \sum_{p=1}^{M(q)} e(q,p)},$$

$H(l)$ would be high when the segment is a harmonic signal. In other words, the degree of voicing for a speech segment can be indicated by the high segmental harmonicity. Figure 5 illustrates the segmental harmonicities of the given utterance. The magnitude of the segmental harmonicity indicates the completeness of the harmonics structure of a segment.

III. SPEECH DISTORTION ANALYSIS

When a speech signal is corrupted by an additive noise, the harmonic structure will be partially distorted. The type and level of corrupting noise determine the degree and distribution of the harmonic structure distortion. Accordingly,
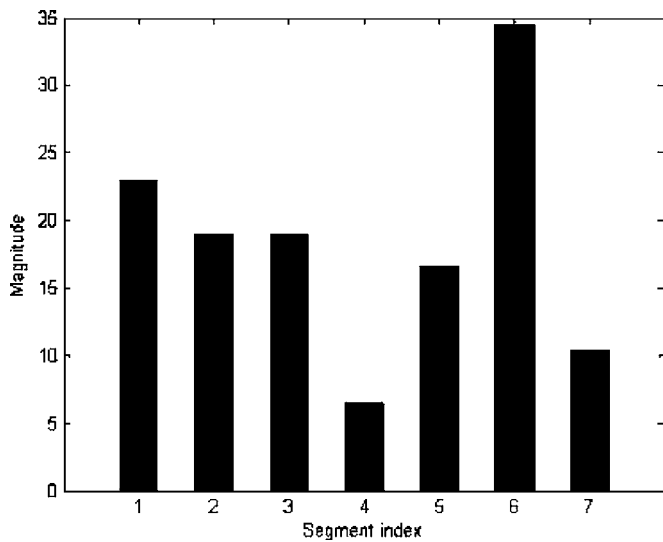
FIG. 5. Segmental harmonicities of an utterance: The sentence is the same as that in Fig. 1.

the spectrum of the speech signal is distorted. The proposed harmonicities can be used to evaluate the speech distortion in different scopes. The grid harmonicity measures the completeness of an individual harmonic in each frame. The segment-spectral harmonicity examines the completeness of an individual harmonic in a speech segment. The temporal harmonicity evaluates the weighted strength of all harmonics in a frame. Finally, the segmental harmonicity is used to give a measure to the entire harmonic structure within a segment. Based on these harmonicity measures, we can acquire the detailed distortion information due to the noise corruption.

Figure 6 shows the case of speech corrupted with car noise at SNR=0 dB. The harmonic structure is partially distorted, but is still easily to be recognized. However, some distortion at lower harmonics is obvious. Many higher harmonics are even invisible in the spectrogram. Besides, the magnitudes of segmental harmonicities are reduced as being compared with the clean speech in Fig. 5. The reduction of segmental harmonicity corresponds to the degree of distortion to a segment.

The examples of segment-spectral harmonicities under noise corruption are illustrated in Fig. 7. It shows that the

additive noise causes different degrees of distortion to harmonic structures. Comparing with the case of clean speech in Fig. 4, the magnitudes of segment-spectral harmonicity for most segments are reduced. More distortion appears at higher harmonics.

IV. ROBUST FUNDAMENTAL FREQUENCY ESTIMATION

The fundamental frequency (F0) estimation is required in many applications, such as speech coding, speech recognition, speech enhancement, and prosody extraction. With such a wide range of interests, many F0 estimation algorithms have been developed to fit their specific applications. A comprehensive study of F0 tracking was presented by Hess (1983). Most F0 estimation algorithms involve four phases: the preparation, the preprocessing, the F0 candidate picking, and the postprocessing. In the preparation phase, the speech signal is windowed and represented in a sequence of frames. The preprocessing stage minimizes those factors that are not helpful or are even harmful to F0 tracking, such as the formant structure. Typically this stage includes some band-pass filtering and nonlinear techniques (Rabiner, 1977). In the stage of F0 candidate picking, F0 candidates are generated via an auxiliary function that assigns higher probability to more likely F0 candidates at each frame. The auxiliary function can be formulated in several domains, such as in time (Qian and Kimaresan, 1996), autocorrelation (Wise *et al.*, 1976), normalized cross correlation (Kasi and Zahorian, 2002) AMDF, cepstrum (Noll, 1967), and subharmonic spectrum (Hess, 1983) domains. The postprocessing stage involves the process of searching for the most possible F0 candidates and forming an F0 track. The most common approach applies the technique of dynamic programming to determine an optimal F0 track.

Though the F0 estimation has been developed for many years, it still remains nontrivial to robustly identify the correct F0 from its double and half values. The uncertainty in picking an F0 candidate makes the F0 estimation algorithms complicated and not robust. However, the validation of F0 candidates based on harmonic structure analysis alleviates this problem. Since a harmonic structure with correct F0 maximizes the temporal harmonicity in a frame, we can
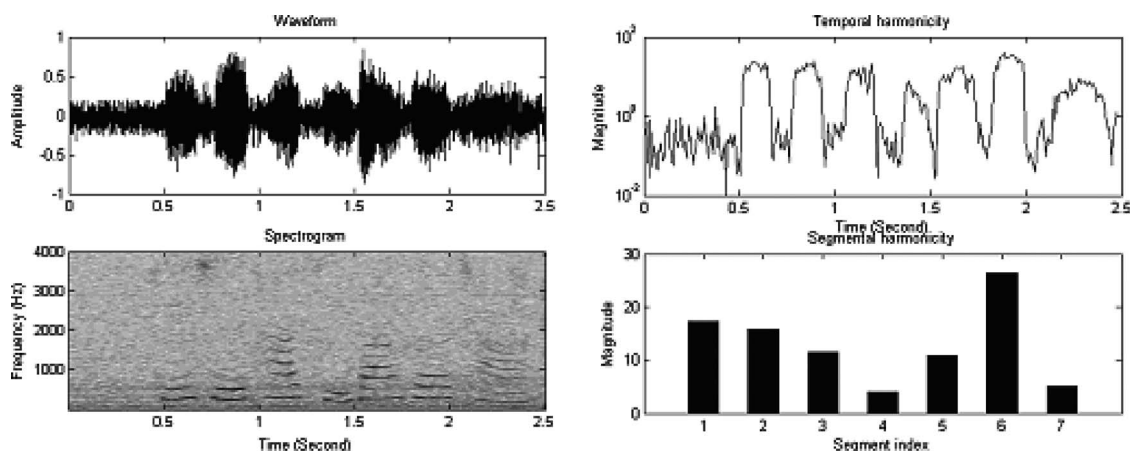
FIG. 6. Temporal harmonicity and segmental harmonicity of a speech corrupted by the car noise: The sentence is the same as that in Fig. 1.

A. Yu and H. Wang: New speech harmonic structure measure
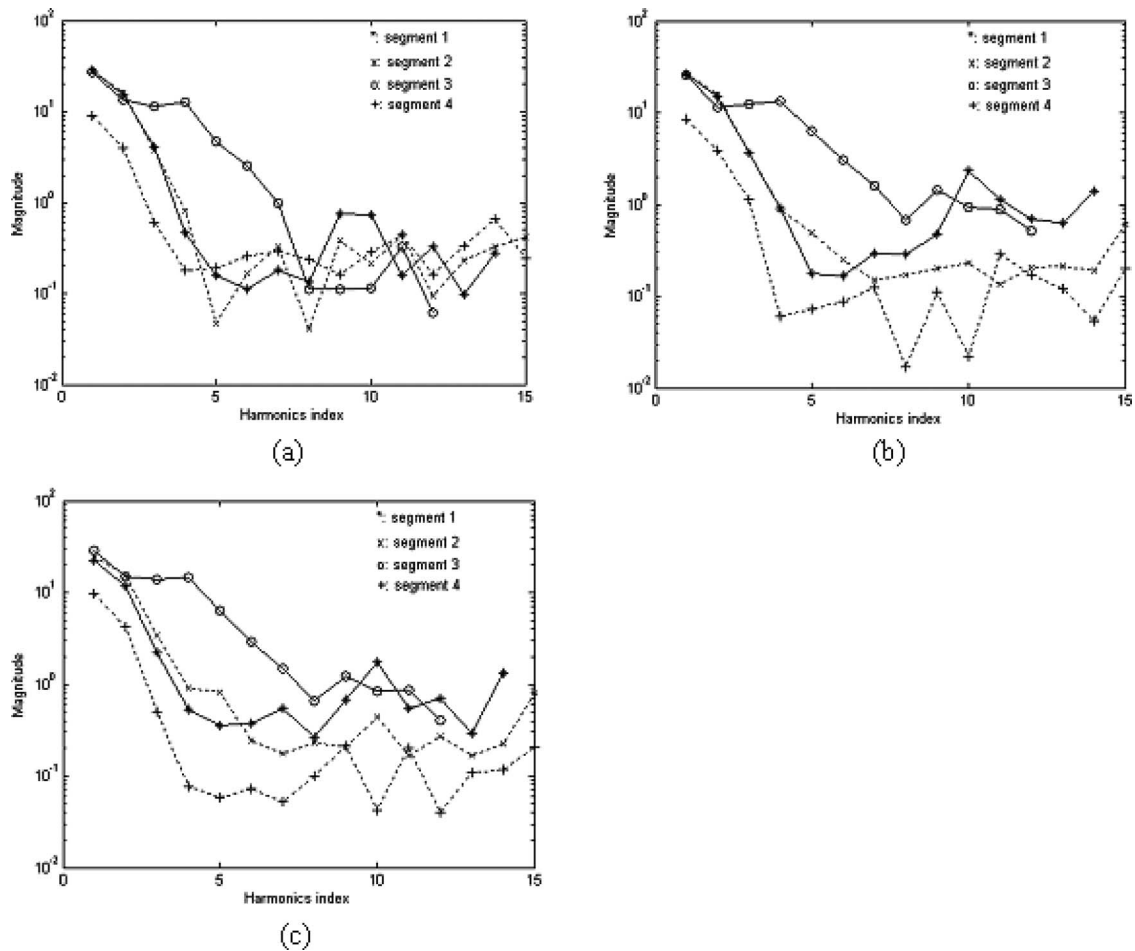
FIG. 7. Segment-spectral harmonicities of first four segments: (a) speech corrupted by white noise, (b) speech corrupted by car noise, and (c) speech corrupted by subway noise. All the noise levels are at SNR = 0 dB..
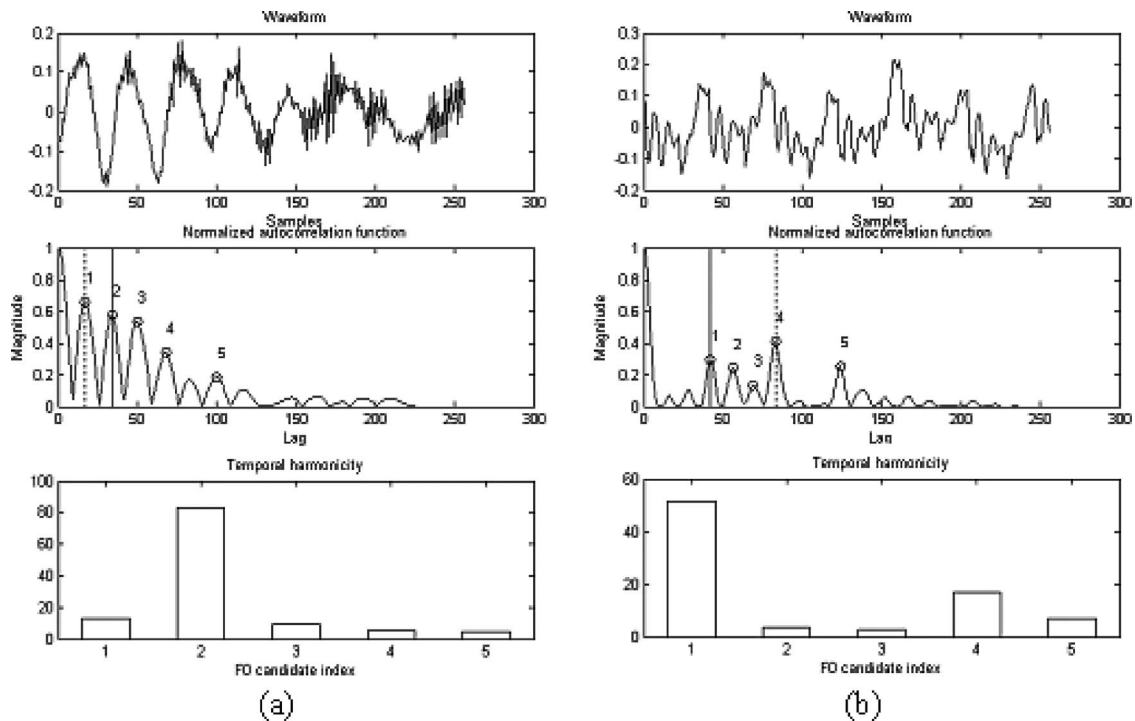
FIG. 8. Harmonicity checking for autocorrelation-based F0 estimation: (a) correction from half estimate error and (b) correction from double estimate error. (Five F0 candidates are marked by small circles. The arrow points to the location of maximal peak where there is the estimated F0 based on autocorrelation. The vertical solid line indicates the corrected F0 after harmonicity checking.)

check the temporal harmonicity to determine the correct F0. Figure 8 displays two examples involving wrong candidates picking frequently encountered in the F0 estimation. The F0 can be identified by observing the waveform of the speech signal and its autocorrelation function. However, the maximal peak of the autocorrelation function may not be the correct F0. Instead, the maximal peak may appear at half or a multiple of the correct F0 so as to lead to an incorrect F0 estimation. Figure 8 shows two examples. First five peaks are indicated by small circles in the plots. If we pick the highest peaks, the example (a) gives a half F0 estimate, and the example (b) gives the double F0 estimate. The temporal harmonicities of five F0 candidates for these two examples are illustrated at the bottom of Fig. 8. According to the temporal harmonicities, we can find that the correct F0 should be at the second peak in example (a) because it gives the highest temporal harmonicity. Similarly, the correct F0 should be at the first peak in example (b). These examples show that the temporal harmonicity can be used for correcting the F0 estimation.

An experiment was conducted to evaluate this validation process. Ten utterances were used as the testing data. Each utterance was corrupted with white noise at levels of SNR ranging from −5 to 30 dB with a 5-dB step. The top five peaks from the autocorrelation function were chosen as F0 candidates. The temporal harmonicity was calculated for each F0 candidate for the validation of estimated F0. No smoothing between frames was performed. A manually labeled F0 track was used as the reference. The F0 estimate was considered to be correct if the estimated F0 was within 20% of the reference F0. The error rate is in the range of 5% at SNR=30 dB to 14% at SNR=−5 dB.

Another experiment was conducted to show the effectiveness of the validation process. The experiment was arranged as follows: (1) The F0 candidates were obtained from five domains, i.e., autocorrelation, AMDF, harmonic peaks, subharmonic sum, and cepstrum. (2) A 3-point median filter was used to filter out the outliers and thus obtained the final F0 track. (3) No harmonicity checking was performed for the baseline testing. (4) The performance was evaluated using

TABLE I. The average temporal harmonicities of noise.

Noise type	Average temporal harmonicity
Airport	2.97
Babble	3.01
Car	1.40
Exhibition	2.15
Restaurant	2.43
Street	2.45
Subway	1.50
Train	2.33
White	1.22
Human speech	25.7

two measures, the gross error and the normalized mean square error. These measures are defined as follows:

(a) Gross errors: The error rate is the percentage of frames with F0 estimate deviating from that of the reference by more than 20%. This measure is based only on voiced frames.

(b) Normalized mean square error: The normalized mean square error is defined as follows;

$$\varepsilon = \frac{\sum [x(i) - y(i)]^2}{\sum [x(i)]^2}, \tag{11}$$

where $x(i)$ denotes the reference track, $y(i)$ represents the computed track, and i is the index of the voiced frame.

The experiment used ten utterances of Mandarin speech. Nine noises listed in Table I were used as additive noises. The speech signals sampled at 8 kHz were degraded by the noises with specific SNR ranging from −10 to 30 dB, with a 5-dB step. Figures 9(a) and 9(b) display the average performances of F0 estimation from five domains. These figures clearly reveal that the F0 estimation performance can be significantly improved by applying the harmonicity checking in the validation process.
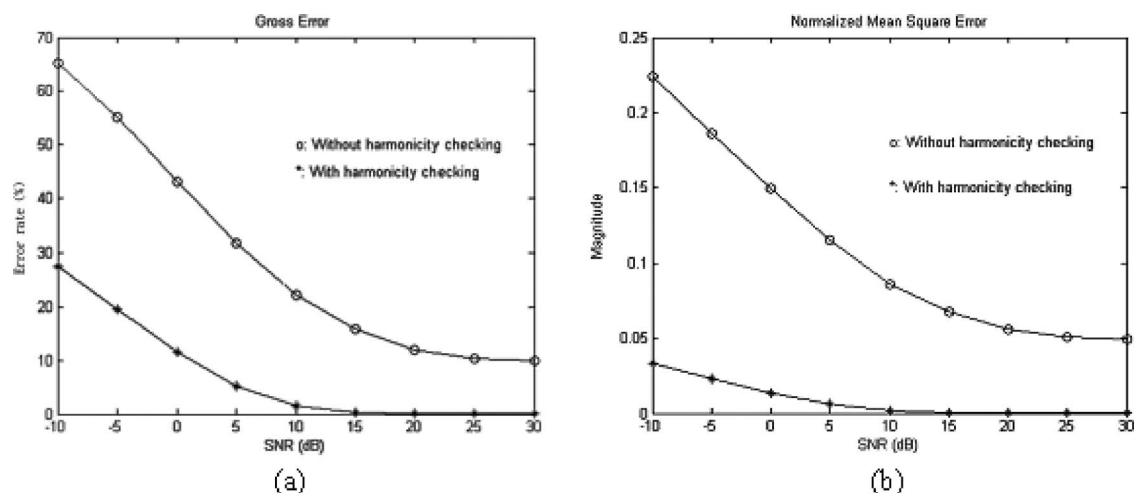
FIG. 9. The performances of F0 estimation: (a) average gross error rates of F0 estimation and (b) average normalized mean square errors of F0 estimation. (The F0 candidates were obtained from five domains, i.e., autocorrelation, AMDF, harmonic peaks, subharmonic sum, and Cepstrum.)

A. Yu and H. Wang: New speech harmonic structure measure
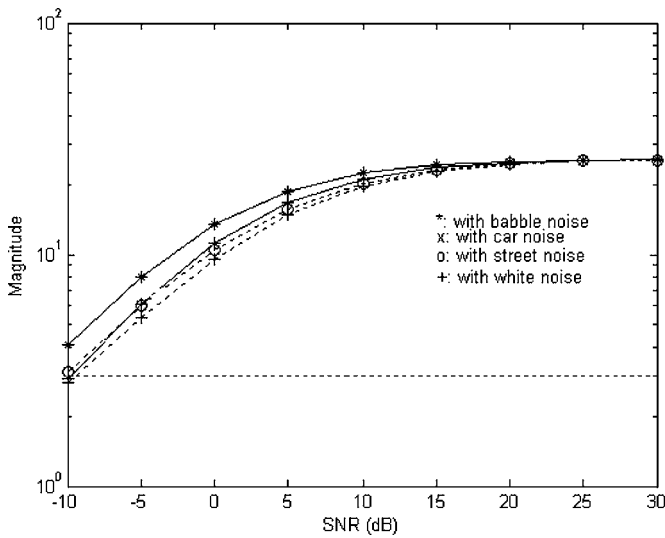
FIG. 10. Average temporal harmonicity of voiced speech evaluated at various noise condition. (The dot line at the magnitude equal to 1.13 indicates the largest average temporal harmonicity of noise, i.e., the babble noise, shown in Table I.)

V. ROBUST VOICING DETECTION

The voicing detection is an important issue in speech processing, such as speech recognition, speech coding, speech enhancement, and telephony applications (Gerven and Xie, 1997; McKinley and Whipple, 1997; Berthommier and Glotin, 1999; Kobatake *et al.*, 1989). The earlier algorithms may be based on Itakura LPC distance measure (Rabiner and Sambur, 1997), energy levels, zero crossing rates (Junqua *et al.*, 1991), cepstral features (Haigh and Mason, 1993), adaptive noise modeling of voice signals (Yoma *et al.*, 1996), and periodicity measurement (Tucker, 1992). Unfortunately, these algorithms are not robust to noise, particularly nonstationary noise. The proposed harmonicity provides another method of voicing detection.

The existence of harmonic structure provides a simple and robust indicator for identifying human speech. Table I
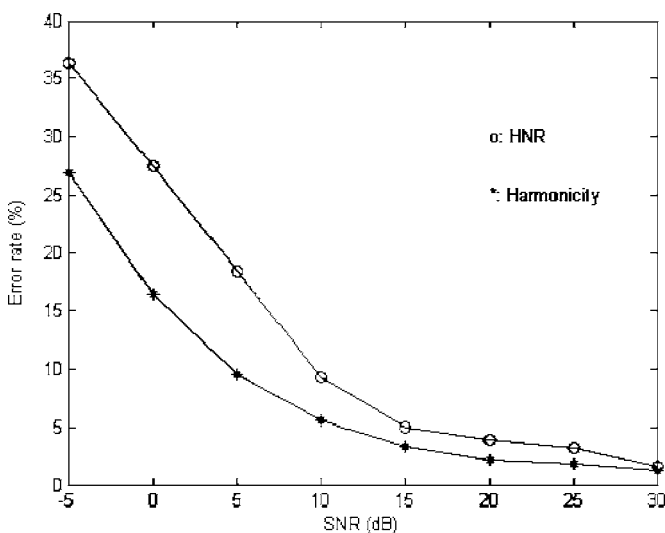
FIG. 11. Comparison of voicing detection methods based on HNR measure and based on harmonic measure.
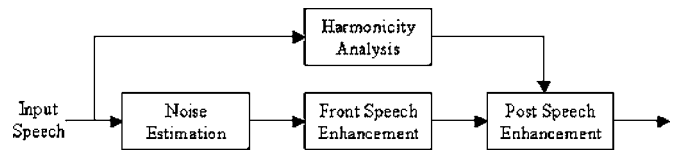
FIG. 12. Structure of the proposed two-stage speech enhancement system.

lists the average temporal harmonicities of nine noises, and the average temporal harmonicity of clean speech. The remarkable difference in temporal harmonicity between noise and speech implies that the temporal harmonicity is an appropriate parameter for voicing discrimination. Based on the information of Table I, it is easy to set a threshold for voicing detection.

In an experiment, the speech was corrupted by various noises. The temporal harmonicities of noisy speech in various noise conditions were computed. Figure 10 shows the average temporal harmonicities of these noisy speech signals. It demonstrated that the temporal harmonicity of noisy speech is still higher than that of noise. This provides a method for robust voicing detection.

The experimental condition for using the temporal harmonicity to determine the voicing status of speech frames is identical to that in the F0 estimation experiment. Figure 11 illustrates the experimental result. The error rates of the conventional method for voicing determination based on harmonics-to-noise ratio (HNR) are plotted for comparison. The error rate is defined as the percentage of frames where the tracker disagrees with the reference in voicing decision. The result reveals that the method based on the average temporal harmonicity outperforms the HNR method.

VI. POSTSPEECH ENHANCEMENT

Many speech enhancement algorithms have been proposed (Lim *et al.*, 1974; Ephraim and Malah, 1984, 1985; Tsoukalas *et al.*, 1997; Wolfe and Godsill, 2000). However, few of these algorithms closely incorporated with speech characteristics. The knowledge of harmonic structures of voiced speech should be incorporated into the enhancement system to further improve the speech quality. Rather than designing a new enhancement algorithm, this study proposes a two-stage scheme: a front enhancement process followed by a postenhancement process. Figure 12 shows the function blocks of this two-stage scheme. An existing enhancement algorithm can be applied as the front enhancement process. The postenhancement process utilizes the detailed speech information obtained from the harmonicity analysis to further improve the speech quality.

For a voiced frame, a comb filter is applied to block out the noise components around the harmonics. The frame-based enhancement often causes discontinuous harmonics. The discontinuity may result in artificial noises. Temporally smoothing the harmonics magnitude in a segment can reduce these unpleasant sounds. The comb filter with gain and bandwidth adapted by the factors derived from temporal and segment-spectral harmonicities is proposed for reducing the artificial noises.
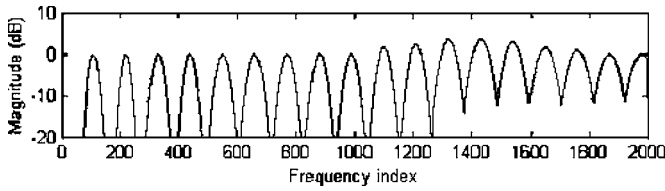
FIG. 13. Spectrum of the comb filter with gain and bandwidth adapted by harmonicities.

The proposed comb filter is implemented in the frequency domain,

$$\Phi(n,k,l) = \sum_{m=1}^{M(n)} g(n,m,l)\,\vartheta(k-mk_0,m,l), \qquad (12)$$

where $M(n)$ denotes the number of harmonics at the nth frame, k_0 is the frequency bin index of the fundamental frequency, $g(n,m,l)$ is a gain function for smoothing the harmonics magnitude in the lth segment, and $\vartheta(k,m,l)$ is the kernel function of the comb filter defined as

$$\vartheta(k,m,l) = e^{-k^2/2\sigma^2(m,l)}, \qquad (13)$$

where $\sigma^2(m,l)$ represents a factor for controlling the bandwidth of the kernel function. Narrower bandwidth of the kernel function makes the shape of the comb filter sharper and hence removes more noise around the harmonics. The sharp comb filter does not produce unnatural sounding for strong harmonics, but incurs artificial sounding for weak ones. To avoid this artificial sounding, $\sigma^2(m,l)$ is formulated as a function inverse proportional to the segment-spectral harmonicity,

$$\sigma^2(m,l) = \frac{\sigma_0^2}{1+\mu H_S(l,m)}. \qquad (14)$$

In Eq. (14), $H_S(l,m)$ is the segment-spectral harmonicity of the mth harmonics in the lth segment. $\sigma_0^2=120$ and $\mu=0.02$ are empirically set for the enhancement experiment.

The gain function $g(n,m,l)$ is formulated as

$$g(n,m,l) = (1-\zeta(m,l)) + \zeta(m,l)\hat{g}(n,m,l), \qquad (15)$$

where $\hat{g}(n,m,l)$ is an adjusted gain function in order to make the harmonics smooth in a segment, and $\zeta(m,l)$ is an interpolation factor defined by

$$\zeta(m,l) = \min\left(\frac{1}{1+e^{c(m)(H_S(l,m)-d(m))}},\lambda(m)\right), \qquad (16)$$

where $c(m)$ and $d(m)$ are factors for controlling the harmonics smoothing for mth harmonics. $d(m)$ is set to 2 and $c(m)$ is set to 0.3 empirically for all m. $\lambda(m)$ is the upper bound for the interpolation factor $\zeta(m,l)$. In this study $\lambda(m)$ is set to 0.2 for all m. The situation of $\zeta(m,l)=0$ corresponds to the case of no harmonics smoothing. $\zeta(m,l)$ is limited to $\lambda(m)$ in order to prevent from oversmoothing. The adjusted gain function $\hat{g}(n,m,l)$ is defined as follows,

$$\hat{g}(n,m,l) = \frac{\hat{e}(n,m,l)}{e(n,m)}, \qquad (17)$$

where $\hat{e}(n,m,l)$ is a target value for obtaining the smooth harmonics,

$$\hat{e}(n,m,l) = \frac{H_T(n)}{\sum_{n=\text{SegBegin}(l)}^{\text{SegEnd}(l)} H_T(n)}\left(\sum_{n=\text{SegBegin}(l)}^{\text{SegEnd}(l)} e(n,m)\right). \qquad (18)$$

Equation (18) indicates that $e(n,m)$ of the mth harmonics is recalculated in the lth segment according to a factor specified by $H_T(n)/\sum_{n=\text{SegBegin}(l)}^{\text{SegEnd}(l)} H_T(n)$. This modification smoothes the discontinuous harmonics caused by the front enhancement process. Figure 13 illustrates a comb filter with adjustable gain and bandwidth.

Subjective and objective evaluations were conducted to evaluate the proposed method. The testing data were ten utterances of Mandarin Chinese. The white noise and F16 noise served as additive noises (Varga *et al.*, 1992). The speech signals sampled at 8 kHz were degraded by the above-mentioned noises, with SNRs ranging from −5 to 10 dB, with a 5-dB step. Three enhancement algorithms, including spectral subtraction, Wiener filtering, and log-
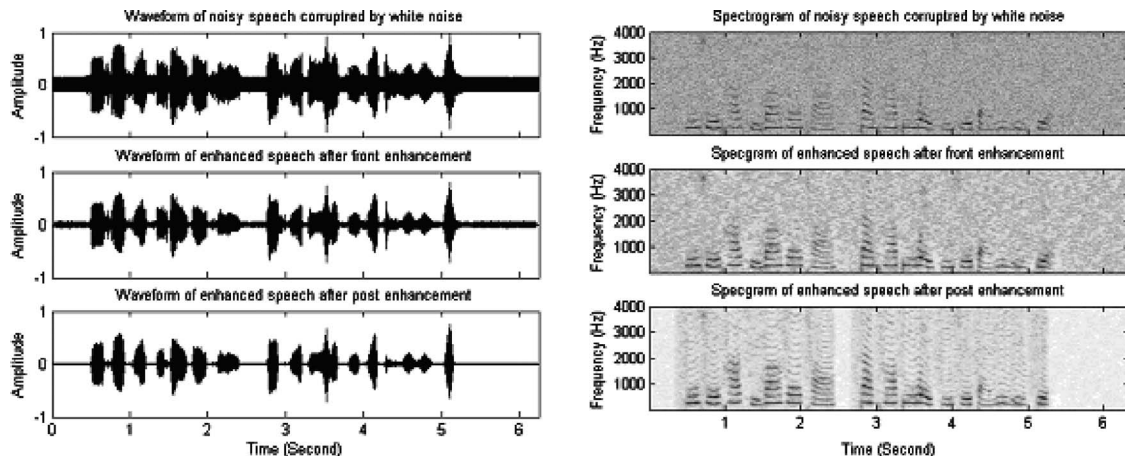
FIG. 14. Result of speech enhancement: The utterance is a sentence spoken in Mandarin Chinese, "qi2 shi2 ta1 xi3 huan1 he1 cha2, dan4 ta1 bing4 wei4 ji2 shi2 fan3 yin4 yi4 jian4. (In fact she likes tea, but she does not response immediately)."

spectral-amplitude (LSA) estimation (Ephraim and Malah, 1985), were separately applied in the front enhancement process.

A minimum statistics method estimated the noise spectrum from the testing utterance for all enhancement algorithms (Martin, 1994). In the postenhancement process, the speech frame contained 256 samples with an overlap of 192 samples. A 2048-point FFT with zero padding transformed the speech waveform into the frequency domain. After the enhancement process in the frequency domain, the spectrum was transformed back to the time domain and truncated to 256 samples. The "overlap-add method" smoothed the speech data between consecutive frames (Oppenheim and Schafer, 1989).

Figure 14 shows the speech enhancement for the case of corrupting by white noise with SNR=5 dB. The given utterance is a sentence spoken in Mandarin Chinese, "qi2 shi2 ta1 xi3 huan1 he1 cha2, dan4 ta1 bing4 wei4 ji2 shi2 fan3 yin4 yi4 jian4 (In fact she likes tea, but she does not response immediately)." Waveforms and spectrograms of the noisy speech and two versions of enhanced speech, with and without the postenhancement process, are displayed.

It is obvious that the noisy speech suffers from a severe distortion in harmonics structures. The existing speech en-

hancement algorithms can reduce the noise. The postenhancement process based on harmonic analysis can further remove the artificial noises. By using the comb filter, a reconstruction of harmonics structures can be achieved. Tables II and III show the segmental SNR improvements with and without the postenhancement process (PE). It indicates that the SNR of enhanced speech in every case is increased and the SNR improvement can be up to 2.8 dB.

The result of MOS evaluation with scores ranging from 1 to 5 is shown in Table IV. Fifteen listeners, including ten males and five females, participated in the listening test. Ten utterances of Mandarin Chinese corrupted with two kinds of noises were played. Table IV shows that the postenhancement process has significantly improved the subjective quality. The improvements mainly come from the fact that the proposed method effectively smoothes the temporal evolvement of harmonics and thereby decreases the occurrence of artificial noises.

VII. CONCLUSION

This paper presents a set of harmonicities, i.e., grid, temporal, segment-spectral, and segmental harmonicities. These new harmonicities can be used for the systematic analysis of

TABLE II. Segmental SNR improvements for speech corrupted by white noise.

Input SNR (dB)	Ephraim-Malah		Spectral subtraction		Wiener filter	
	Without PE (dB)	With PE (dB)	Without PE (dB)	With PE (dB)	Without PE (dB)	With PE (dB)
−5	9.5	11.3	9.5	10.7	10.2	11.2
0	8.0	10.4	7.3	9.8	8.4	10.1
5	6.3	8.9	5.1	7.9	6.5	9.0
10	4.6	6.5	3.0	5.4	4.6	6.3

PE stands for post enhancement process.

TABLE III. Segmental SNR improvements for speech corrupted by F16 noise.

Input SNR (dB)	Ephraim-Malah		Spectral subtraction		Wiener filter	
	Without PE (dB)	With PE (dB)	Without PE (dB)	With PE (dB)	Without PE (dB)	With PE (dB)
−5	8.2	10.6	7.9	9.6	8.3	10.1
0	6.9	9.3	5.9	8.3	7.2	8.9
5	4.1	7.0	3.8	6.2	4.5	6.8
10	2.7	4.8	1.3	3.5	3.3	4.7

TABLE IV. Comparison of MOS for various speech enhancement methods.

Input SNR (dB)	Noisy speech	Ephraim-Malah		Spectral subtraction		Wiener filter	
		Without PE	With PE	Without PE	With PE	Without PE	With PE
−5	1.24	2.30	2.97	2.27	2.63	2.48	2.87
0	1.58	2.96	3.32	2.84	3.05	2.94	3.20
5	2.15	3.25	3.89	3.05	3.62	3.16	3.73
10	2.63	4.14	4.37	3.93	4.13	4.02	4.21

harmonic structure in an utterance so as to be applied to many tasks in speech and harmonic signal processing. In this study, applications to the speech distortion analysis, the robust F0 estimation, the robust voicing detection, and the speech enhancement are demonstrated. A series of experiments was conducted to show the effectiveness of the proposed harmonics analysis. It allows a detailed understanding of the spectral distortion in noise corruption conditions. The temporal harmonicity can be used for the validation of F0 estimation and the robust voicing detection. When the temporal and segment-spectral harmonicities are used to control an adjustable comb filter in the postenhancement process of a two-stage speech enhancement system, the artificial noise can be further reduced.

APPENDIX: LIST OF PARAMETERS

Parameter	Related equation	Description
$e(n,m)$	(1)	Sum of weighted magnitude spectral densities around the mth harmonics at the frequency bin index mk_0 in nth frame
$[(m-\eta)k_0,(m+\eta)k_0]$	(1)	Summation range for computing $e(n,m)$
$S(n,k)$	(1)	Magnitude spectral density of an analyzed signal in nth frame and kth frequency bin index
$\phi(k)$	(1), (2)	Harmonics selecting function
$h_D(n,m)$	(3)	Local spectral dominance of mth harmonics evaluated in nth frame
$D_T(n,m)$	(4)	Spectral magnitude variation between adjacent frames of mth harmonics evaluated in nth frame
$h_T(n,m)$	(5)	Temporal variation factor of mth harmonics evaluated in nth frame, with value between 0 and 1
$h_N(n,m)$	(6)	Neighbor dominance of mth harmonics evaluated in nth frame
$H_G(n,m)$	(7)	Grid harmonicity of mth harmonics evaluated in nth frame
$H_T(n)$	(8)	Temporal harmonicity evaluated in nth frame
$\alpha(n,m)$	(8)	Weighting factor for computing $H_T(n)$
$H_S(l,m)$	(9)	Segment-spectral harmonicity of mth harmonics evaluated over lth segment
$\beta(n,m,l)$	(9)	Weighting factor for computing $H_S(l,m)$
$H(l)$	(10)	Whole harmonic measure within lth segment
$\gamma(n,m,l)$	(10)	Weighting factor for computing $H(l)$
$\Phi(n,k,l)$	(12)	Comb filter for kth harmonics in nth frame and lth segment
$g(n,m,l)$	(12)	Gain function for smoothing the harmonics magnitude at mth harmonics
$\vartheta(k,m,l)$	(12), (13)	Kernel function of the comb filter
$g(n,m,l)$	(15)	Gain function

A. Yu and H. Wang: New speech harmonic structure measure

Parameter	Related equation	Description
$\hat{g}(n,m,l)$	(15), (17)	Adjusted gain function
$\zeta(m,l)$	(15), (16)	Interpolation factor for computing $g(n,m,l)$
$\hat{e}(n,m,l)$	(17), (18)	Target value for obtaining the smooth harmonics

Arifianto, D., and Kobayashi, T. (**2003**). "IFAS-based voiced/unvoiced classification of speech signal," *Proc. IEEE-ICASSP*, Vol. **I**, pp. 812–815.

Berthommier, F., and Glotin, H. (**1999**). "A measure of speech and pitch reliability from voicing," in *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, edited by F. Klassner, Computational Auditory Scene Analysis (CASA) workshop, pp. 61–70.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," *Proceedings of the Institute of Phonetic Sciences* **17**, pp. 97–110.

de Cheveigné, A., and Kawahara, H. (**2001**). "Comparative evaluation of $F0$ estimation algorithms," in *Proc. European Conference on Speech Communication and Technology. Eurospeech 2001*.

Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-32**, 1109–1121.

Ephraim, Y., and Malah, D. (**1985**). "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-33**, 443–445.

Gerven, S. V., and Xie, F. (1997). "A comparative study of speech detection methods," Eurospeech-97, pp. 1095–1098.

Haigh, J. A., and Mason, J. S. (**1993**). "Robust voice activity detection using cepstral features," in *Proc. IEEE TENCON*, China, pp. 321–324.

Hess, W. (**1983**). "*Pitch determination of speech signals*," in Springer Series of Information Sciences (Springer-Verlag, Berlin).

Junqua, J. C., Reaves, B., and Mak, B. (**1991**). "A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognize," in *Proc. Eurospeech*, pp. 1371–1374.

Kasi, K., and Zahorian, S. A. (**2002**). "Yet another algorithm for pitch tracking," *Proc. IEEE-ICASSP* (IEEE, New York).

Kobatake, H., Tawa, K., and Ishida, A. (**1989**). "Speech/non-speech discrimination for speech recognition system under real life noise environments," *Proc. IEEE-ICASSP* (IEEE, New York), p. 365.

Krom, G. de (**1993**). "A cepstrum based technique for determining an harmonics-to-noise ratio in speech signals," J. Speech Hear. Res. **36**, 254–265.

Lim, J. S., Oppenheim, A. V., and Braida, L. D. (**1974**). "Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition," in *Speech enhancement*, Alan V. Oppenheim series, edited by J. S. Lim (Prentice-Hall, Englewood Cliffs, NJ), pp. 88–92.

Martin, R. (**1994**). "Spectral subtraction based on minimum statistics," *Proc. of the Seventh European Signal Processing Conference, EUSIPCO-94*, September, pp. 1182–1185.

McAulay, R. J., and Quatieri, T. F. (**1986**). "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**(4), pp. 744–754.

McKinley, B. L., and Whipple, G. H. (**1997**). "Model based speech pause detection," *Proc. IEEE-ICASSP* (IEEE, New York), pp. 1179–1182.

Murphy, P. J. (**2000**). "A cepstrum-based harmonics-to-noise ratio in voice signals," *Proc. ICSLP*, Vol. **4**, pp. 672–675.

Nehorai, A., and Porat, B. (**1986**). "Adaptive comb filtering for harmonic signal enhancement," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**(5), 1124–1138.

Noll, A. M. (**1967**). "Cepstrum pitch determination," J. Acoust. Soc. Am. **14**, 293–309.

Oppenheim, A. V., and Schafer, R. W. (**1989**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Qi, Y., and Hillman, R. E. (**1997**). "Temporal and spectral estimations of harmonics-to-noise ratio human voice signals," J. Acoust. Soc. Am. **102**(1), 537–543.

Qian, X., and Kimaresan, R. (**1996**). "A variable frame pitch estimator and test results," *Proc. IEEE-ICASSP*, Atlanta, GA, Vol. **1**, pp. 228–231.

Rabiner, L. R. (**1977**). "On the use of autocorrelation analysis for pitch detection," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-25**, 24–33.

Rabiner, L. R., and Sambur, M. R. (**1997**). "Voiced-unvoiced-silence detection using the Itakura LPC distance measure," *Proc. IEEE- ICASSP*, pp. 323–326.

Tsoukalas, E., Mourjopoulos, J., and Kokkinakis, G. (**1997**). "Speech enhancement based on audible noise suppression," IEEE Trans. Speech Audio Process. **5**(6), 497–514.

Tucker, R. (**1992**). "Vad using a periodicity measure," *IEE Proceedings, Communications, Speech and Vision*, Vol. **139**, No. 4, pp. 377–380.

Varga, A., Steenneken, H. J. M., Tomlinson, M., and Jones, D. (1992). "The NOISEX-92 study on the effect of additive noise on automatic speech recognition," documentation included in the NOISEX-92 CD-ROMs.

Wise, J. D., Caprio, J. R., and Parks, T. W. (**1976**). "Maximum-likelihood pitch estimation," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-24**, 418–423.

Wolfe, J., and Godsill, S. (**2000**). "Towards a perceptually optimal spectral amplitude estimator for audio signal enhancement," *Proc. IEEE-ICASSP*, Vol. **2**, pp. 821–824.

Yoma, N. B., McInnes, F., and Jack, M. (**1996**). "Robust speech pulse-detection using adaptive noise modeling," Electron. Lett. **32**, 1350–1352.

Using image processing to detect and classify narrow-band cricket and frog calls

T. Scott Brandes[a)]
Tropical Ecology Assessment and Monitoring (TEAM) Initiative, Conservation International,
1919 M Street, NW, Washington, D.C. 20036

Piotr Naskrecki
Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge,
Massachusetts 02138

Harold K. Figueroa
Bioacoustics Research Program, Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road,
Ithaca, New York 14850

An automatic call recognition (ACR) process is described that uses image processing techniques on spectrogram images to detect and classify constant-frequency cricket and frog calls recorded amidst a background of evening sounds found in a lowland Costa Rican rainforest. This process involves using image blur filters along with thresholding filters to isolate likely calling events. Features of these events, notably the event's central frequency, duration and bandwidth, along with the type of blur filter applied, are used with a Bayesian classifier to make identifications of the different calls. Of the 22 distinct sonotypes (calls presumed to be species-specific) recorded in the study site, 17 of them were recorded in high enough numbers to both train and test the classifier. The classifier approaches 100% true-positive accuracy for these 17 sonotypes, but also has a high false-negative rate (over 50% for 4 sonotypes). The very high true-positive accuracy of this process enables its use for monitoring singing crickets (and some frog species) in tropical forests. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2355479]

I. INTRODUCTION

Biodiversity documentation is the first step in its effective conservation. Yet practical implementation of species-level surveys in tropical ecosystems invariably runs into the problem of taxonomic impediment, or the lack of expertise to provide the necessary identification of collected organisms. In addition, the need to physically collect specimens to produce species diversity estimates often proves costly, time consuming, and logistically difficult. For some groups of organisms using their acoustic signals provides a viable alternative to specimen collecting, and allows for monitoring of species richness, presence or absence of indicator species, or even estimates of abundance.[1]

Acoustic signaling is a widespread behavior among many terrestrial and aquatic animals.[2] Cicadas, orthopteroid insects (crickets, katydids, and relatives), frogs, birds, and bats are only some of the groups that can be monitored acoustically. Crickets (Orthoptera: Grylloidea) are particularly well suited for the purpose of acoustic monitoring because the range of frequencies employed by most species as a whole (1–12 kHz) is audible to humans, their signals are simple (few species employ frequency modulation), and their calls are species-specific.[3]

In most species of crickets the males produce audible calls using a modified area on their first pair of wings known as the stridulatory apparatus. Its basic design consists of a row of teeth on the underside of the right wing (the stridulatory file) that produce sounds when rubbed against a hardened edge (the scraper) of the left wing. The signal is amplified by the vibration of a membranous area of the wing (the mirror). In some species additional sound amplification and redirection is achieved by using leaves as acoustic baffles, or by singing from horn-shaped burrows. The resulting signal serves primarily as an attractant to conspecific females, and in most cases is unique and easily recognizable. This allows us to record and recognize individual acoustic signatures of each species (sonotypes) present in an area. The specificity of cricket calls even allows for using them to describe species in the absence of distinctive morphological characters typically used by insect taxonomists,[4,5] and to use the calls to identify species in the field.[6]

In many tropical ecosystems crickets dominate the acoustic space both in terms of the number of species signaling at any given moment, and the frequency spectrum of their calls. Most of the cricket calls' energy is within the audible part of the sound spectrum, and only some members of the subfamily Phalangopsinae produce ultrasonic calls in the range above 22 kHz.[7] This concentration of crickets' calls within the audible part of the spectrum, and their abundance in most ecosystems, makes them ideal candidates for long-term monitoring. At the same time crickets are very sensitive indicators of environmental change, as exemplified recently by the extinction of a number of Hawaiian species

―――――――――――――
[a)]Electronic mail: s.brandes@conservation.org
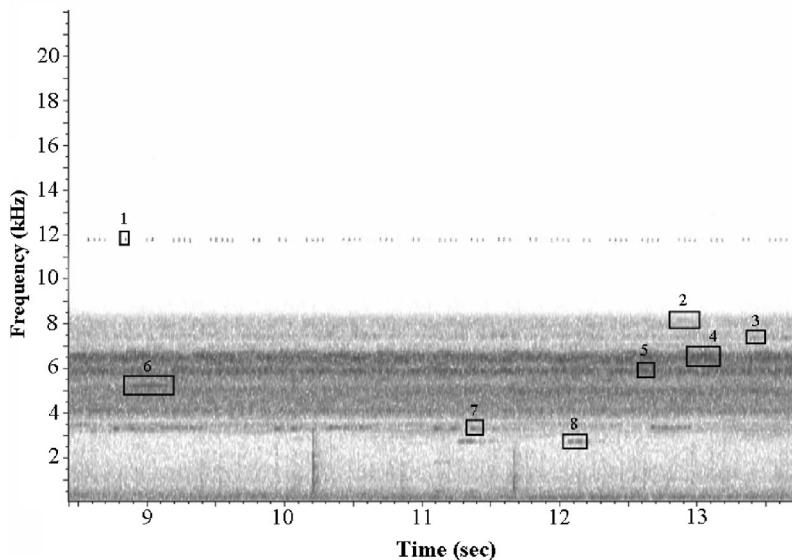
FIG. 1. Typical spectrogram of canopy recording. At least 8 sonotypes of crickets (1–6) and frogs (7, 8) present.

due to invasion of exotic ant species.[8] Acoustic monitoring has already been successfully implemented in detection of rare or endangered insect species in temperate zones,[9] and the same technique can be used to monitor species richness and the presence of species indicative of particular changing habitat parameters in tropical ecosystems.[10]

Multiple encouraging papers on automated recognition of bioacoustics signals have been recently published for a range of animals, including bats,[11–13] birds,[14,15] anurans,[16] and Orthoptera.[17–19] Since cricket songs have a simple structure, the more involved techniques used in human speech recognition that have shown promise in bird song identification,[15] such as dynamic time warping and hidden Markov models, add unnecessary complication. Current research on Orthoptera call recognition[17–19] rely heavily on time-domain signal processing and require a high signal to noise ratio in order to properly analyze call waveforms. This same problem holds true with frequency-domain features commonly used in sound recognition such as spectrum flux and spectral power,[20,21] and mixed species choruses that vary in composition make these tools less useful. This reliance on low background noise severely limits the ability of these approaches to accurately identify crickets in tropical rainforests since numerous cricket species often call simultaneously and some of the quieter sounds present might be from crickets we are trying to detect. In the work presented here, an image processing approach is suggested as a method for detecting and classifying cricket (and some frog) calls recorded amidst the constellation of background noises typically found shortly after sunset in tropical forests.

II. MATERIALS AND METHODS

A. Sound recordings

In the work presented here we used field recordings of crickets from a lowland rainforest in La Selva Biological Station located in NE Costa Rica. Some of the recordings were made with in-house autonomous recording units that were placed midcanopy, and deployed to acoustically monitor crickets nightly for weeks at a time. However, to quicker

assemble a library of cricket calls at the site, hand-held recorders were also used. The autonomous recorders were placed midcanopy to effectively record sounds originating from both the canopy and the ground.[22] These devices used a PDA to make a 10 min recording nightly, starting 1 h after local sunset. PDAs where chosen since software can be written for them to make high-quality digital recordings at scheduled intervals. We used Hewlett Packard iPAC 5555 PDAs with a custombuilt power supply (rechargeable NiMH battery pack) and weatherproof case, and in-house scheduled-recording software.[23] The devices record sound at 44.1 kHz with 16-bit resolution and use a 16-element microphone and amplifier designed by the Bioacoustics Research Program at Cornell University. The microphone has a toroid beam-pattern and is very effective at collecting sounds within the canopy.

A spectrogram of a typical canopy recording is shown in Fig. 1; most of the sounds are from crickets and frogs. There are at least 8 distinct sonotypes (6 crickets and 2 frogs) present amidst a background of more distant individuals. While the frog calls are known for La Selva, no known libraries of cricket calls exist for the area. Work is still ongoing in assigning species labels to each acoustic signature and in the work presented here we designate our classifications as sonotype number.

B. Automatic call recognition (ACR)

The ACR process we use involves operating on a spectrogram image of a sound. Spectrograms created with a FFT size of 512 points, a Hanning window, and a 50% window overlap provide sufficient resolution in both time (5.8 ms) and frequency (86.1 Hz) to allow detection and classification of dozens of sonotypes at our site. As shown in Fig. 2, once
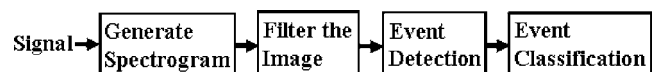
FIG. 2. Steps in automatic call recognition using the image processing presented here.
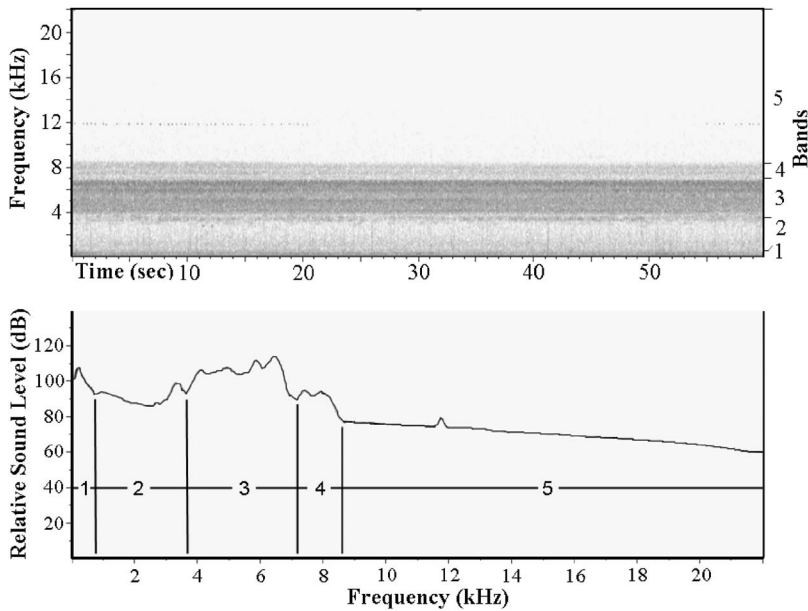
FIG. 3. 1 min average of spectrogram and zones to look for cricket and frog calls. Zone 1=0–0.75 kHz; Zone 2=0.75–3.6 kHz; Zone 3=3.6–7.1 kHz; Zone 4 =7.1–8.6 kHz; Zone 5=8.6–22 kHz.

the spectrogram is generated, the image is then filtered, events are detected, and finally the events are classified. Details for each of these steps follow.

C. Image filtering

The images are filtered in 3 steps. Initially the images are broken into 4–5 frequency bands within which to look for calls. The image is then sent through a blur filter. Finally, the image is sent through a thresholding filter, where a threshold level is applied independently for each frequency band.

The images are broken into 4–5 frequency bands based on relative background noise level. This allows the cutoff threshold of the thresholding filter to be adjusted for each frequency band, allowing calls in other frequency bands that are fainter than the background in the loudest frequency band to stand out. These frequency bands are chosen by calculating the average value of the intensity within the first minute of the spectrogram across all frequencies (Fig. 3). The different frequency bands are chosen by setting their boundaries as the positions which bound the various frequency intensity peaks. Bands are set to group intensity peaks such that intensity minimums are not higher than relative peaks within the band. In Fig. 3, band 3 (3.6 kHz–7.1 kHz) has the most intensity. If it were merged with frequency band 2 or 4, calls in those frequency ranges would not have a high enough dB level to be above the cutoff level of the thresholding filter and would not be detected. In the work presented here, these bands were chosen manually, however this step could be safely automated by using the frequency bin values generated in the first minute of each recording since they stay consistent throughout each of our 10 min canopy recordings.

1. Blur filters

For many loud constant-frequency calls, this filter can be bypassed. However, when the signal to noise ratio drops, bypassing this filter and going straight to the thresholding filter will extract only fragments of full calls. This is shown

in the long call centered around 5.25 kHz in Fig. 4, where only the first 80% of the call is shown as continuous when the thresholding filter is applied. With a higher signal to noise ratio, the entire call would be revealed as continuous. The event detection algorithm discussed shortly looks for continuous call chunks, and smaller call fragments would not necessarily be grouped together. A remedy for this is to use a filter to blur the image in order to get call fragments to connect. This works very well with a boxcar averaging blur filter applied prior to the thresholding filter, set to slightly blur only in time, not in frequency. With a boxcar averaging filter, the grey-scale value of each pixel in the new image $\widetilde{P}_{r,c}$ in row r and column c is the average of all the initial image pixel values $P_{r,c}$ within a rectangular swath (mask) centered over it, with width w and height h. With indices i and j, the boxcar averaging filter can be written as

$$\widetilde{P}_{r,c} = \frac{\sum_{i=-h}^{h} \sum_{j=-w}^{w} P_{r+i,c+j}}{(2h+1)(2w+1)}.$$

In the work presented here, we set $w=1$ and $h=0$, creating a slight blur in time, and none in frequency. This is effective for connecting most cricket call fragments. With our sampling rate and spectrogram parameters, this limits our time resolution to be on the order of 10 ms, which is sufficiently long for the crickets we recorded. The results of both the boxcar averaging blur filter with the thresholding filter and the thresholding filter on its own are shown in Fig. 5. The top image is of the spectrogram without any filtering, the middle with a thresholding filter, and the bottom image with both a boxcar averaging blur and thresholding filter. The thresholding filter is applied on the second frequency band. Two calls of the same sonotype are shown. The leftmost call is from a nearby individual, and the rightmost is from a more distant individual. This demonstrates how the signal detection degrades as the signal to noise ratio diminishes. Notice how the rightmost call in
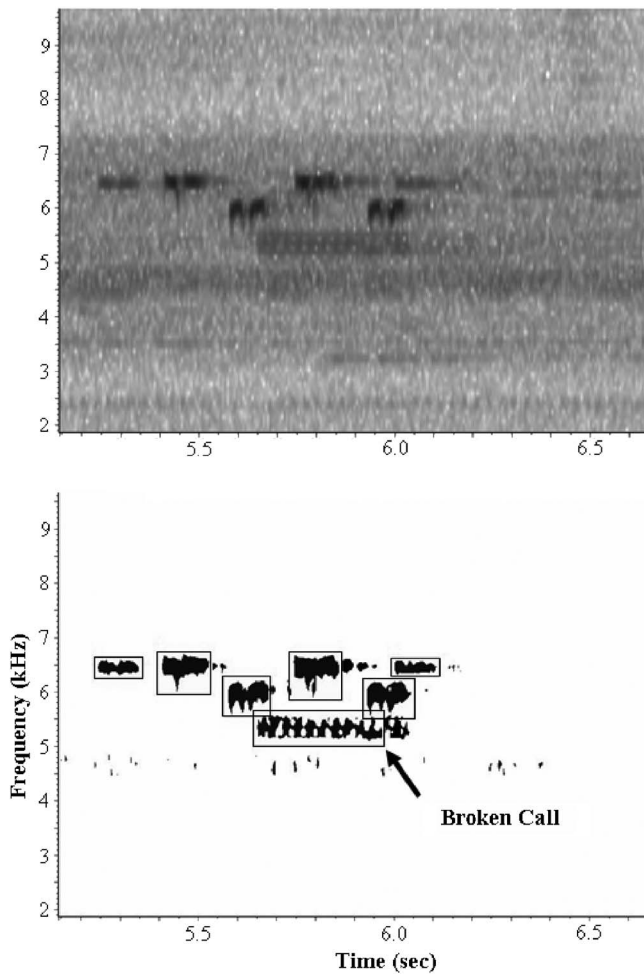
FIG. 4. Thresholding filter example. Calls are separated from the background. Note that the lower frequency call (5.5 kHz) in the filtered image (bottom) becomes fragmented toward the end. The louder calls present remain intact in the filtered image. Blurring the image before applying the thresholding filter minimizes call fragmentation.

FIG. 5. Boxcar averaging blur filter with thresholding filter example. The thresholding filter is used in the bottom 2 images, between 750 Hz–3.6 kHz. The boxcar averaging blur filter is applied in the lower image. Note how the call fragments of the rightmost call in the middle image are connected to form a continuous call when both filters are applied in the bottom image, allowing easier identification. The call on the left is from an individual of the same species that is closer to the microphone.

the thresholding only filter is fragmented, whereas the call is continuous with the boxcar averaging blur filter.

Some cricket calls consist of a series of discrete chirps and this pulse train is repeated periodically. In these cases, applying a Gaussian blur filter

$$\widetilde{P}_{r,c} = \frac{\sum_{i=-2h}^{2h} \sum_{j=-2w}^{2w} P_{r+i,c+j} e^{-(i^2/2h^2)} e^{-(j^2/2w^2)}}{(4h+1)(4w+1)}$$

can smooth the pulse train into a continuous dash and facilitate detection. Here, the pixel values of the new image are generated on a weighted average (based on a normalized Gaussian distribution) of the pixel values of the initial image within a particular swath centered over each pixel. This can be seen in Fig. 6 with the sonotype consisting of a pulse train near 3 kHz. Images on the left are without the blur filter and the ones on the right are made with a Gaussian blur filter with $w=3$ and $h=0.5$. The lower 2 images show how the detection results change once the thresholding filter is applied. In the bottom left image where the blur filter is not used, each individual pulse is detected, whereas in the bottom right image where the blur filter is used, the entire

pulse train is detected as a single unit. In this image it is also interesting to see how the 5.5 kHz sonotype stands out more prominently from the background as the image is blurred. In the work presented here, we chose to use both of these blur filters empirically, based on their performance, and in the detection process, they are each run independently on all the sounds and are used to detect different sonotypes. In this way, the type of blur filter used serves as one of the features or parameters with which to classify a sonotype.

2. Thresholding filter

The purpose of the thresholding filter is to set a threshold of an intensity level required to consider a call present, and to condition the image for the event detection process. This process is done for each frequency band independently, each with a slightly different threshold level. The threshold level for each frequency band is set as an intensity level relative to the sound present in that frequency band (based on a running 10 s average). For the sound shown in Fig. 3, the sound level is a little higher in bands 3 and 4 than in bands 2 and 5, and a threshold level of 8 dB relative to the within-frequency band sound level is used, whereas a 9 dB threshold is used in bands 2 and 5. No known sonotypes from the site are low enough in frequency to be in band 1, so a threshold level on that band was not set. Adjusting the threshold levels slightly depending on the signal to noise ratio allows the calls to stand out better; however, keeping the threshold level at 8 dB would have produced good results. For the work presented here, these threshold levels were chosen manually, and better results might be obtainable
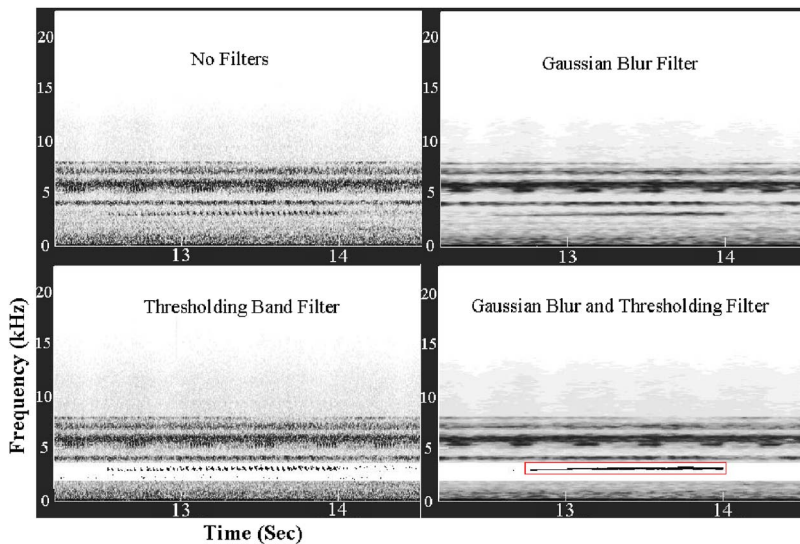
FIG. 6. Gaussian-blur with thresholding filter example. The blur filter is applied to the images on the right. In the lower images, a thresholding filter is applied. Note that when both filters are applied, the entire pulse train is selected as the event, not just the individual pulses as in the bottom left image. This facilitates the call recognition process. Also notice how the sonotype at 5.5 kHz is more prominent in the blurred image. This shows how blurring can make a pattern more pronounced.

in an automated process by running the same sections at a range of threshold levels and using the results that yield the most events.

D. Event detection

After the filtering, a block detection algorithm is run to find events within the spectrogram that might be cricket calls. This algorithm goes through the filtered image keeping track of blocks of nonblank pixels that meet specific structural requirements based on 4 parameters: maximum bandwidth (8 pixels, 689 Hz), maximum time allowed to exceed bandwidth (4 pixels, 23 ms), minimum duration (3 pixels, 17 ms), and maximum duration (241 pixels, 1.4 s). The values for these parameters are set near the extreme values of the collection of sonotypes to minimize computation time on events that will not be classified as known sonotypes. During the exploratory phase, when sonotype ranges are being determined, this added efficiency can lead to missed calls. Since we are looking for constant-frequency calls, there is an additional constraint that as the image is swept along the time axis, at least one pixel in the current time band must fit within the frequency bandwidth swept out to that point by the event. This allows the event bandwidth to increase throughout a call, but does not allow it to capture steep sweeping calls; though, it could be modified to do so. The "maximum time allowed to exceed bandwidth" parameter allows short-time broadband noise such as a rain drop to not interfere with detecting a call.

E. Event classification

Once the events in the spectrogram are detected, features are automatically measured from these pixel clusters and the calls are classified as particular sonotypes, or as unknown events. The features measured from each event are the central frequency, bandwidth, duration, and type of blur filter used. These features separate the sonotypes recorded at our site with little overlap, and we designed a Bayesian classifier to label events as belonging to particular sonotypes. Distributions of the features of the training set of sonotypes are used to generate 95% confidence intervals for these features

for each sonotype, and these intervals serve as the decision bounds for the classifier. For the sonotypes we recorded enough times to both train and test, these 4 features completely separate each sonotype. However, for 2 of the sonotypes that we recorded in too low of number to both train and test the classifier, the additional features of the time gap between similar calls as well as the number of similar calls in a pulse train are required to fully separate these sonotypes from others. Since we have very few recordings of these sonotypes, these 2 additional features were not automatically generated in our testing sets, and these additional sonotypes went undetected in our testing trials (none were even detected manually).

III. FIELD RESULTS

Twenty-two distinct constant-frequency sonotypes were identified from the site. Two sonotypes, ST101 and ST102, are from frogs, whereas the remaining sonotypes are from crickets. Most of the sonotypes are separable with the central frequency, duration, and blur filter used. Figure 7 shows the distribution of training samples for the 22 sonotypes recorded, plotted as central frequency vs duration. A Gaussian blur filter is used to find sonotype 2; each of the other 21 sonotypes are found by applying a boxcar averaging blur filter. A 95% confidence interval of the distribution of each sonotype is shown as an ellipse over the distribution points. A few of the ellipses have some overlap (ST 15, ST 22, ST 23) and require additional features for separation. These 95% confidence intervals serve as the decision boundaries for the Bayesian classifier. Only 1 call for sonotype 8 and 9 was found, so the arbitrarily set confidence interval around them only serves to show the approximate bounds with which to try to detect those calls (no additional calls were found for either sonotype).

A confusion matrix for the 17 sonotypes recorded enough times to both train and test, separated with only 4 features (central frequency, duration, bandwidth, and blur filter), is shown in Table I. Sonotypes 6, 8, 9, 14, and 15, present in Fig. 7, were not found enough times to do both training and testing, and are absent from Table I. "Unknown"
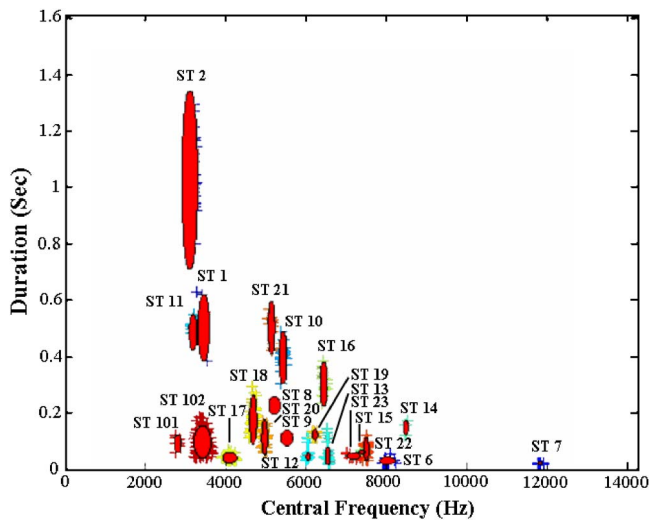
FIG. 7. (Color online) Two of the features used to separate sonotypes, central frequency and call duration, are shown for 22 sonotypes. The estimated 95% confidence interval around the distributions are represented by ellipses. Gaps between the sonotypes indicate possible openings for additional species we did not record.

TABLE II. Accuracy measures for 17 sonotypes based on the confusion matrix in Table I.

	True positive	False negative	Train/test detections (N)
ST 1	97.78%	78.85%	3/44
ST 2	100%	17.86%	25/46
ST 7	100%	71.55%	290/865
ST 10	100%	25.00%	20/114
ST 11	100%	26.32%	8/28
ST 12	100%	17.85%	50/336
ST 13	98.15%	24.64%	75/159
ST 16	100%	67.82%	24/65
ST 17	100%	11.58%	298/741
ST 18	100%	42.78%	56/226
ST 19	100%	38.89%	15/55
ST 20	100%	08.62%	40/371
ST 21	100%	41.30%	12/27
ST 22	100%	08.33%	50/363
ST 23	100%	37.50%	10/35
ST 101	100%	68.42%	7/6
ST 102	100%	42.16%	279/834

calls are defined as events that have feature values outside of the estimated ranges of the sonotypes defined by the training set. Since many crickets call in pulse trains, instead of choosing training and testing sets randomly, entire pulse trains were used for either training or testing. This was done by designating the first samples of sonotypes found within sound files manually as the training set, and sonotypes found manually in subsequent files as the testing set. Accuracy percentages from the confusion matrix are shown in Table II, along with the number of training and testing samples. Sample number of test detections are counted as the number of correct detections of a sonotype during the testing phase.

IV. DISCUSSION

One of the interesting aspects of the sonotypes in Fig. 7 is their distribution. The degree to which the sonotype fea-

ture values overlap provides insight into how accurate the classifier is. With no overlap in these distributions, the classifier will be 100% accurate in its classifications. The shape of the collective distribution of sonotype features and the degree to which this area is filled with sonotypes is also noteworthy. The sonotype feature distributions collectively in Fig. 7 appear bound by a curve, limiting cricket calls between approximately 3 kHz and 12 kHz, and allowing a wider range of call durations at the lower frequencies, gradually shortening maximum call duration up to the 12 kHz calls. It is likely that the bounds of this curve are determined by the stridulatory apparatus of the crickets at our site. If the area of usable feature space is bound by cricket physiology, then the degree to which this area is saturated with cricket sonotypes might be insightful in determining the degree to which the ensemble of cricket species at a site have remained

TABLE I. Confusion matrix for 17 sonotypes. Actual classifications are in the leftmost column, and the assigned classifications in the topmost row. Correct identifications are indicated on the diagonal. Events labeled as "unknown" by the classifier are shown in the column labeled "UNKN."

	UNKN	ST1	ST2	ST7	ST10	ST11	ST12	ST13	ST16	ST17	ST18	ST19	ST20	ST21	ST22	ST23	ST101	ST102
UNKN	32	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST1	164	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST2	10	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST7	2175	0	0	865	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ST10	38	0	0	0	114	0	0	0	0	0	0	0	0	0	0	0	0	0
ST11	10	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0
ST12	73	0	0	0	0	0	336	0	0	0	0	0	0	0	0	0	0	0
ST13	52	0	0	0	0	0	0	159	0	0	0	0	0	0	0	0	0	0
ST16	134	0	0	0	0	0	0	3	65	0	0	0	0	0	0	0	0	0
ST17	97	0	0	0	0	0	0	0	0	741	0	0	0	0	0	0	0	0
ST18	169	0	0	0	0	0	0	0	0	0	226	0	0	0	0	0	0	0
ST19	35	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0	0	0
ST20	35	0	0	0	0	0	0	0	0	0	0	0	371	0	0	0	0	0
ST21	19	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0
ST22	33	0	0	0	0	0	0	0	0	0	0	0	0	0	363	0	0	0
ST23	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0
ST101	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0
ST102	608	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	834

intact. Gaps in feature space could also be due to lack of time spent collecting recordings for an area, and sonotype overlaps could occur without causing interference problems if those species seldom call at the same time of day or year. Interesting gaps that exist in Fig. 7 include the frequency range between 8.6 kHz and 11.7 kHz, and calls over 0.3 s long between 3.6 kHz and 5 kHz. Implications for these gaps require further study.

For most sonotypes shown in Fig. 7, the duration varies more than central frequency. Some of this is an accurate representation for the variation in call length, such as for sonotype 16. For other sonotypes, such as sonotype 2, the lower end of the distribution is a little misleadingly stretched. When the sonotypes are recorded with a high signal to noise ratio, their features are measured accurately. However, when the sonotype is fainter, only fragments of the call are detected. The blur filter goes a long way towards connecting these fragments, but with sonotype 2 in particular, the intensity of the call increases throughout the duration of the call (Fig. 6) and the measured duration of the call will vary depending on how loud the quieter beginning of the call is (how far away the insect is from the microphone). By accepting a larger range in the call duration, the Bayesian classifier is able to detect more distant individuals than it would otherwise. By increasing the acceptable range of a feature like this, the possibility of misclassification increases if the range of features of different sonotypes begin to overlap. To avoid misclassification in this situation, additional features are needed. Sonotype 2 is the only sonotype listed that is measured after a Gaussian blur is applied, so the use of the filter itself would safely allow accurate classification if sonotype 2 had feature distributions that overlapped different sonotypes.

The central frequency of the sonotypes have some spread to them, some more than others. The variation captured in Fig. 7 mostly reflects the variation within a pulse train and the variation between individuals. Air temperature can play a large role in central frequency shift of a sonotype but did not for the work presented here since the air temperature during the recordings varied little within the extremes of 22.5 °C and 24.4 °C. In cases where air temperature can change more dramatically, training set sonotypes need to be recorded at multiple temperatures. Central frequency does not noticeably shift as a function of air temperature for all cricket species, but for the species that air temperature does affect this way, the shift in central frequency is linear.[24,25] For these sonotypes, once the linear relationship between central frequency and air temperature is found, air temperature during a recording should be used as a parameter to estimate the acceptable distribution of central frequencies for these sonotypes when trying to detect them.

A. Improving performance

The approach to ACR described in this work very accurately identifies our sonotypes, but the emphasis is clearly on keeping a high true-positive accuracy at the expense of false-negative accuracy. Our application is in monitoring sonotypes from canopy recordings, and since most crickets call in

pulse trains, presence and absence can be readily determined even by detecting only some of the pulses in those pulse trains. In general, there is no need to detect each call within the entire sequence of calls. Two separate sources of error arose to lower the true-positive accuracy in 2 sonotypes down from 100%, although still above 97.5% (Table II). In one instance, a consecutive string of sonotype 102 calls completely overlapped to create an event that matched sonotype 1. In the other instance, a few faint calls of sonotype 16 had degraded to the point that they resembled the shorter and smaller bandwidth calls of sonotype 13. Each of these instances point out how the duration and bandwidth features detected from events can lead to some misclassifications with sonotypes that have overlapping central frequency distributions. To decrease these misclassifications, additional features such as the time gap between similar events and the number of similar events within a specified amount of time (duration of a pulse train) can be used to make accurate identifications. These features look to have little variance for many of the sonotypes, particularly the short duration ones which are more likely to lead to misclassifications. To detect these features, most of the calls within a sequence of calls will need to be detected. Call sequences recorded with a high signal to noise ratio have enough of their calls detected to measure these features. Requiring these features for classification would increase the false-negative percentage for these sonotypes, but would also improve the true-positive percentage to nearly 100%.

Most crickets call in a rather continuous pulse train, and when trying to establish the presence or absence of sonotypes, it is not necessary to detect each call within a pulse train, only some of them. Table II shows that the false-negative classification percentage of 4 sonotypes is above 50%. While not generally a problem when detecting presence or absence of a sonotype, it is desirable for these percentages to be lower. The reason for these high values is mostly due to not having a wide enough distribution of feature values for those sonotypes. For sonotypes 1 and 101, this is likely due to having small training sets ($N=3$ and 7, respectfully). The other 2 sonotypes, ST7 and ST16, have many more training samples, but they are likely not varied enough. Each training set was chosen as the first sonotype instances encountered, not randomized from a group of known instances. Randomizing this training process would likely allow the feature distributions for these sonotypes to be expanded while keeping the number of training samples the same.

B. Finding novel calls

When looking at class distributions similar to the ones shown in Fig. 7, the possibility of using clustering techniques to determine class boundaries in an automated way looks intriguing. In this paper, the class assignment of the training sonotypes is done manually by grouping calls that look nearly identical and are often part of a continuous series of call sequences, and they are presumed to come from the same species. If the training was done with isolated calls recorded with a high signal to noise ratio, using automated

clustering techniques might well be possible. The training sets, though, are created from field recordings with various background sounds. This results in miscellaneous noises being classified as events along with sonotypes. Since the sonotypes are not necessarily more numerous than miscellaneous noises, the events are manually classified.

After the training phase, clustering of unknown calls might provide an automated way to find novel calls, at least those that have a duration longer than 0.2 s. Most of the unknown events are short in duration, and applying a minimum duration filter goes far towards weeding out events that are not novel sonotypes. Once enough of these filtered events are found, a clustering method such as from Diday[26] might be useful in finding clusters that correspond to novel sonotypes. A dynamic clustering method such as this would likely be a good choice since it allows the clusters to be elongated, as many of the manually classified sonotype clusters are.

V. CONCLUSIONS

The ACR process described in this paper detects constant-frequency cricket and frog calls with a high true-positive accuracy. The accuracy of this approach largely depends on how comprehensively sonotypes are known for an area, and to what extent the sonotype feature values overlap. The success of this process shows that image processing techniques applied to spectrogram images can provide a useful approach in ACR applications. This process works with recordings made in tropical forests, where multiple species can call simultaneously and numerous background noises are present, and can be used to monitor presence or absence of cricket species that have constant-frequency calls, or relative abundance measures if there are multiple recording locations. Potential for misclassification of cricket sonotypes is largely dependent on how complete the libraries of cricket calls for an area are, and how well variation in those calls is known. The more complete an understanding there is for the type of calls that can be encountered, the more accurately the feature-space can be partitioned for each sonotype, and the more accurately the automated process will run.

[1] P. F. Fischer, U. Schulz, H. Schubert, P. Knapp, and M. Schmöger, "Quantitative assessment of grassland quality: Acoustic determination of population sizes of Orthopteran indicators species," Ecol. Appl. **7**, 909–920 (1997).

[2] W. J. Bailey, *Acoustic Behaviour of Insects. An Evolutionary Perspective* (Chapman and Hall, London, 1991).

[3] K. Riede, "Acoustic monitoring of Orthoptera and its potential for conservation," Journal of Insect Conservation **2**, 217–223 (1998).

[4] T. J. Walker, "Cryptic species among sound-producing ensiferan Orthoptera (Gryllidae and Tettigoniidae)," Q. Rev. Biol. **39**, 345–355 (1964).

[5] D. Otte, "The crickets of Hawaii: Origin, systematics and evolution," Orthopterists' Society, Philadelphia, pp. 1–396 (1994).

[6] D. R. Ragge and W. J. Reynolds, *The Songs of the Grasshoppers and Crickets of Western Europe* (Harley, Colchester, 1998).

[7] L. Desutter-Grandcolas (personal communication).

[8] D. Otte (personal communication).

[9] C. Froehlich and E. Holtzem, "Eine neue Methode zum Auffinden und Bestimmen von Heuschrecken (Saltatoria) im Freiland," Z. angew. Zool. **74**, 501–503 (1987).

[10] F. Nischk and K. Riede, "Bioacoustics of two cloud forest ecosystems in Ecuador compared to lowland rainforest with special emphasis on singing cricket species," in *Epiphytes and Canopy Fauna of the Otongan Rain Forest (Ecuador). Results of the Bonn-Quito Epiphyte Project, Funded by the Volkswagen Foundation*, edited by J. Nieder and W. Barthlott (2001), Vol. **2**, pp. 217–242.

[11] N. Vaughan, G. Jones, and S. Harris, "Identification of British bat species by multivariate analysis of echolocation call parameters," Bioacoustics 7, 189–207 (1997).

[12] S. Parsons and G. Jones, "Acoustic identification of 12 species of echolocating bats by discriminant function analysis and artificial neural networks," J. Exp. Biol. **203**, 2641–2656 (2000).

[13] S. Parsons, "Identification of New Zealand bats in flight from analysis of echolocation call by artificial neural networks," J. Zool. **253**, 447–456 (2001).

[14] S. E. Anderson, A. S. Dave, and D. Margoliash, "Template-based automatic recognition of birdsong syllables from continuous recordings," J. Acoust. Soc. Am. **100**, 1209–1219 (1996).

[15] J. A. Kogan and D. Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," J. Acoust. Soc. Am. **103**, 2185–2196 (1998).

[16] A. Taylor, G. Watson, G. Grigg, and H. McCallum, "Monitoring frog communities: An application of machine learning," in Proceedings of the 8th Innovative Applications of Artificial Intelligence Conference (1996), pp. 1564–1596.

[17] E. D. Chesmore, O. P. Femminella, and M. D. Swarbrick, "Automated analysis of insect sounds using time-encoded signals and expert systems-A new method for species identification," in *Information Technology, Plant Pathology and Biodiversity*, edited by P. Bridge, P. Jeffries, D. R. Morse, and P. R. Scott (CAB International, Wallinford, 1998), pp. 273–287.

[18] E. D. Chesmore, "Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals," Appl. Acoust. **62**, 1359–1374 (2001).

[19] C. Dietrich, G. Palm, K. Riede, and F. Schwenker, "Classification of bioacoustic time series based on the combination of global and local decisions," Int. J. Pattern Recognit. Artif. Intell. **37**, 2293–2305 (2004).

[20] C. Xu, N. C. Maddage, and X. Shao, "Automatic music classification and summarization," IEEE Trans. Speech Audio Process. **13**, 441–450 (2005).

[21] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall, New Jersey, 1993).

[22] K. Riede, "Bioacoustic monitoring of insect communities in a Bornean rain forest canopy," in *Canopy Arthropods*, edited by N. E. Stork, J. Adis, and R. K. Didham (Chapman and Hall, London, 1997), pp. 442–452.

[23] T. S. Brandes, "Acoustic Monitoring Protocol," Tropical Ecology Assessment and Monitoring (TEAM) Initiative set of biodiversity monitoring protocols, Center for Applied Biodiversity Science, Conservation International (2005), http://www.teaminitiative.org (navigate to "Protocols") (last visited September 2006).

[24] T. J. Walker, "Factors responsible for intraspecific variation in the calling songs of crickets," Evolution (Lawrence, Kans.) **16**, 407–428 (1962).

[25] H. C. Gerhardt and F. Huber, *Acoustic Communication in Insects and Anurans, Common Problems and Diverse Solutions* (University of Chicago Press, Chicago, 2002).

[26] E. Diday, "The dynamic clusters method in nonhierarchical clustering," Int. J. Comput. Inf. Sci. **2**, 61–68 (1973).

Detection of acoustic cavitation in the heart with microbubble contrast agents *in vivo*: A mechanism for ultrasound-induced arrhythmias

Claudio Rota, Carol H. Raeman, Sally Z. Child, and Diane Dalecki[a]
Biomedical Engineering and the Rochester Center for Biomedical Ultrasound, University of Rochester,
309 Hopeman Building, P.O. Box 270168, Rochester, New York 14627

Ultrasound fields can produce premature cardiac contractions under appropriate exposure conditions. The pressure threshold for ultrasound-induced premature contractions is significantly lowered when microbubble contrast agents are present in the vasculature. The objective of this study was to measure directly ultrasound-induced cavitation in the murine heart *in vivo* and correlate the occurrence of cavitation with the production of premature cardiac contractions. A passive cavitation detection technique was used to quantify cavitation activity in the heart. Experiments were performed with anesthetized, adult mice given intravenous injections of either a contrast agent (Optison®) or saline. Murine hearts were exposed to ultrasound pulses (200 kHz, 1 ms, 0.1–0.25 MPa). Premature beats were produced in mice injected with Optison® and the likelihood of producing a premature beat increased with increasing pressure amplitude. Similarly, cavitation was detected in mice injected with Optison® and the amplitude of the passive cavitation detector signal increased with increasing exposure amplitude. Furthermore, there was a direct correlation between the extent of cavitation and the likelihood of ultrasound producing a premature beat. Neither premature beats nor cavitation activity were observed in animals injected with saline and exposed to ultrasound. These results are consistent with acoustic cavitation as a mechanism for this bioeffect. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2346132]

I. INTRODUCTION

Alterations in cardiac rhythm can occur in response to ultrasound at appropriate exposure conditions. Early work by Smailys *et al.* (1981) demonstrated that continuous-wave (CW) ultrasound could stop ventricular fibrillation in the dog heart *in vivo*. Lithotripter fields can produce premature cardiac contractions in mammalian and amphibian hearts *in vivo* (Dalecki *et al.*, 1991; Delius *et al.*, 1994). The threshold for producing a premature contraction with a lithotripter is 2–5 MPa when the pulse is delivered to the heart during the most sensitive phase of the cardiac cycle (i.e., TP segment, diastole) (Dalecki *et al.*, 1991; Delius *et al.*, 1994). A single pulse of ultrasound at high amplitude can also produce premature contractions in laboratory animals (Dalecki *et al.*, 1993a, b; MacRobbie *et al.*, 1997). The threshold for producing a premature contraction in the murine heart with a single 5-ms pulse of ultrasound at 1.2 MHz is ~2 MPa (MacRobbie *et al.*, 1997). The threshold increases with increasing frequency and decreasing pulse duration (Dalecki *et al.*, 1993a, b). A series of repetitive, high-amplitude ultrasound pulses can capture and pace the heart *in vivo* (Dalecki *et al.*, 1993a). Studies have shown that neither thermal nor radiation force mechanisms are responsible for ultrasound-induced premature contractions (Dalecki *et al.*, 1993b, 1997a).

Ultrasound contrast agents are stabilized microbubbles that are injected intravascularly to enhance diagnostic ultrasound imaging. Some ultrasound imaging techniques that use contrast agents depend upon the acoustic excitation of the microbubbles *in vivo*. Acoustic cavitation describes the interaction of a sound field with an existing gas cavity. Under appropriate exposure conditions, cavitation can produce various physical phenomena, including localized shear stresses, high collapse temperatures and pressures, microstreaming, and shock formation (e.g., Young, 1989). Evidence from numerous studies indicates that the presence of microbubble contrast agents in the blood can decrease the threshold for various ultrasound-induced biological effects *in vitro* and *in vivo*, such as hemolysis, capillary rupture, and sonoporation (for review see AIUM, 2000; NCRP, 2002; Dalecki, 2005a).

The stimulation of a premature cardiac contraction without contrast agents requires a high-amplitude ultrasound pulse of long duration, as discussed above. Thus, there is no evidence for the production of premature cardiac contractions with diagnostic ultrasound in the absence of contrast agents. However, recent studies indicate that when microbubble contrast agents are present in the blood, the threshold for ultrasound-induced premature contractions is reduced significantly compared to the threshold in the absence of contrast agents (van der Wouw *et al.*, 2000; Li *et al.*, 2003, 2004; Dalecki *et al.*, 2005b; Rota, 2005). For instance, the acoustic pressure threshold for producing a premature beat in the mouse with a 1.2-MHz, 5-ms pulse of ultrasound is approximately an order of magnitude lower with contrast

---

[a]Author to whom all correspondence should be addressed. Electronic mail: dalecki@bme.rochester.edu

agents (Dalecki *et al.*, 2005b) compared to the threshold without contrast agents (MacRobbie *et al.*, 1997). Furthermore, premature contractions can be produced with diagnostically relevant ultrasound exposures when contrast agents are in the blood (van der Wouw *et al.*, 2000; Li *et al.*, 2003, 2004; Dalecki *et al.*, 2005b; Rota, 2005). Li *et al.* (2003, 2004) found a threshold of ~1 MPa for the production of premature contractions in rats injected with a microbubble contrast agent and exposed to a commercial diagnostic imaging device operating at 1.7 MHz. The threshold for premature contractions in mice injected with contrast agent and exposed to a single, 10-μs pulse of ultrasound at 1.2 MHz is also ~1 MPa (Dalecki *et al.*, 2005b). Premature beats have also been reported in humans injected with an experimental contrast agent and exposed to diagnostic ultrasound at 1.66 MHz with end-systolic triggering and a mechanical index (MI) of 1.5 (van der Wouw *et al.*, 2000). The increased sensitivity of the heart to ultrasound exposure when contrast agents are present is consistent with acoustic cavitation as a mechanism for ultrasound-induced premature cardiac contractions.

In this study, we tested directly the role of acoustic cavitation as a mechanism for ultrasound-induced premature contractions in the presence of microbubble contrast agents. Experiments were designed to measure ultrasound-induced cavitation in the murine heart *in vivo* and to determine whether the occurrence of cavitation is correlated with the production of premature cardiac contractions. A passive cavitation detection (PCD) technique was used to quantify cavitation activity in the heart. A low ultrasound frequency (200 kHz) and long pulse duration (1 ms) were used to lower the threshold for acoustic cavitation. We hypothesized a direct correlation between the occurrence of cavitation and the production of premature cardiac contractions *in vivo*.

II. EXPERIMENTAL METHODS

A. Exposure apparatus

The acoustic source for this investigation was a 200-kHz, 4-cm diameter, unfocused transducer (Model P19, Airmar Technology Corp., Milford, NH). A wave-form generator was used to produce a sinusoidal signal (Hewlett Packard, Model HP33120) at 200 kHz with pulse duration of 1 ms. An attenuator (Kay Elemetrics Corp., PMAX) and rf amplifier (Model 1040L, ENI, Rochester, NY) controlled the amplitude of the signal delivered to the transducer. Acoustic fields were calibrated daily with a probe hydrophone (TC-4038, Reson Inc., Goleta, CA) in a tank of degassed and deionized water. The hydrophone was attached to and moved by a three-way positioning system. The transaxial −6-dB beam width of the transducer was 3.0 cm at a distance of 9 cm from the source.

The apparatus used in this study is shown in Fig. 1. This top view of the water tank illustrates the relative orientation of the 200-kHz source and the 5-MHz passive cavitation detection (PCD) transducer with respect to the exposure site (vial or mouse). The 5-MHz PCD transducer was positioned at a 45-deg angle with respect to the 200-kHz source. For *in vitro* studies, the exposure site consisted of a 4.6-mL dispos-

TOP-VIEW
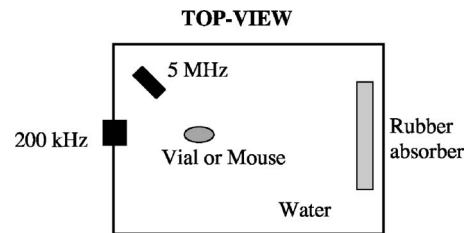
FIG. 1. Exposure setup for passive cavitation detection (PCD) *in vitro* (with vials) and *in vivo* (with mice). The ultrasound source was a 200-kHz unfocused transducer. The PCD transducer was a 5-MHz focused source.

able polyethylene pipet (Fisher Scientific) positioned at a distance of 9 cm, on-axis, from the 200-kHz source. For *in vivo* studies, mouse hearts were also positioned at a distance of 9 cm, on-axis, from the 200-kHz transducer.

A 5-MHz focused transducer (V309, Panametrics, MA) was used as the PCD transducer. For this transducer, the −3-dB axial beam width was 3.5 cm and the −3-dB transaxial beam width was 1 mm. Alignment of the focus of the 5-MHz PCD transducer with the exposure site of the 200-kHz source was accomplished by using the PCD transducer as a pulser receiver (Model 5052PR, Panametrics, MA). A hydrophone was first used to determine coordinates from a fixed pointer in the water tank to the exposure site in the 200-kHz field. Using these coordinates, a 0.1-in. diameter steel ball was then positioned at the exposure site. The 5-MHz PCD transducer was focused on the exposure site by maximizing the received echo signal from the steel ball.

The output of the 5-MHz PCD transducer was connected to a 2-MHz high-pass filter (Model 3940, Krohn-Hite Corp., MA). The cut-off frequency of 2 MHz ensured removal of the 200-kHz incident and scattered field from the spectrum of the PCD output. The filtered PCD output signal was displayed on a digital oscilloscope (Model LT342 WaveRunner, LeCroy Corp., NY). The extent of cavitation activity was quantified as the root-mean-square (rms) of the signal received by the PCD transducer (e.g., Everbach *et al.*, 1997).

B. *In vitro* protocol

Experiments were performed initially with an *in vitro* model in order to validate the cavitation detection technique and provide insight for interpreting later measurements of cavitation *in vivo*. Experiments *in vitro* were performed using cylindrical disposable vials (i.e., polyethylene transfer pipets, Fisher Scientific) approximately 1.4 cm in diameter and 4 cm in height. Each vial was filled with either (1) air, (2) degassed water, or (3) degassed water with Optison®. In samples with microbubble contrast agent, the Optison® was administered as a single injection (0.02 mL) into the pipet immediately prior to the experiment. Each vial was positioned on axis, at a distance of 9 cm from the 200-kHz source, using coordinates obtained by hydrophone measurements from a fixed pointer in the tank to the exposure site.

The PCD apparatus was used to quantify the extent of cavitation activity produced in the vial by exposure to pulses of ultrasound from the 200-kHz source. For each vial (air, water, and water+Optison®), ten, 200-kHz ultrasound pulses

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Rota *et al.*: Detection of cavitation in the heart   2959

were delivered with a pulse repetition period of ~2 s. The duration of each pulse was 1 ms. Peak pressure amplitudes investigated were 0.05, 0.1, 0.2, and 0.25 MPa. Each vial was exposed to only one pressure amplitude (total of ten pulses per vial). During each exposure, the PCD output was recorded and downloaded to a computer. Measurements were also performed without any vial in the field (free field).

C. *In vivo* protocol

All animal protocols were approved by the University of Rochester Committee on Animal Resources. Adult, male (CD1) mice (~30 g) were anesthetized by intramuscular injections of ketamine hydrochloride (60 mg/kg) and xylazine hydrochloride (10 mg/kg) using a 26-gauge needle. Following anesthesia, the ventral chest area was shaved and depilated (Nair®, Church & Dwight Co., Inc., Princeton, NJ) to minimize entrapment of air bubbles on the skin during exposure. Each mouse was secured on a custom-made holder and supported by limb restraints. Dissection of several mice determined the location of the mouse heart to be ~1-cm cephalad to the caudal tip of the sternum and 2–3 mm to the animal's left side. This point was marked with a dot on the skin surface to aid in positioning the heart in the acoustic field. The animal holder was then attached to the three-way positioning system and placed in the water tank. Using coordinates between a fixed pointer and the exposure site determined using the hydrophone, the marked dot on the skin was positioned at the exposure site. The focus of the PCD detector was aligned with the exposure site of the 200-kHz source as described previously. The PCD transducer was then used as a pulse/echo device to ensure that the focus was on the heart and not the highly reflective lungs.

A three-lead electrocardiogram (ECG) was obtained by subcutaneous placement of 25-gauge needles in the limbs. The ECG signal was acquired via a data acquisition system (PowerLab/8SP, ML785, ADInstruments, Grand Junction, CO) that allowed direct monitoring of the ECG on a personal computer and storage of the ECG for further analysis. A front-end preamplifier acquired the ECG signal from the animal (BioAmplifier ML136, ADInstruments, Grand Junction, CO). The ECG signal was sampled at a rate of 1 kHz. A custom-designed triggering device was used to trigger the 200-kHz source at the desired time during the cardiac cycle (diastole). Both the ECG and triggering signal were recorded by the data acquisition system.

Experiments were performed with mice injected with either Optison® or saline. Injections were performed with a 25-gauge needle from a modified ButterflyOptison® infusion set (Dalecki *et al.*, 1997b) inserted into the tail vein of the mouse, and secured with surgical tape (Micropore™, 3M, St. Paul, MN). Each animal received two bolus injections of either Optison® or saline at separate times during the experiment. For Optison® injections, 0.01 mL of the contrast agent was diluted in 0.3 mL of saline to obtain a dose of 3 μL/bolus. Each animal was administered with the same total injection volume (100 μL/bolus). The first bolus was delivered prior to the first ultrasound pulse, and the second bolus was given halfway through the experiment. Mice in-
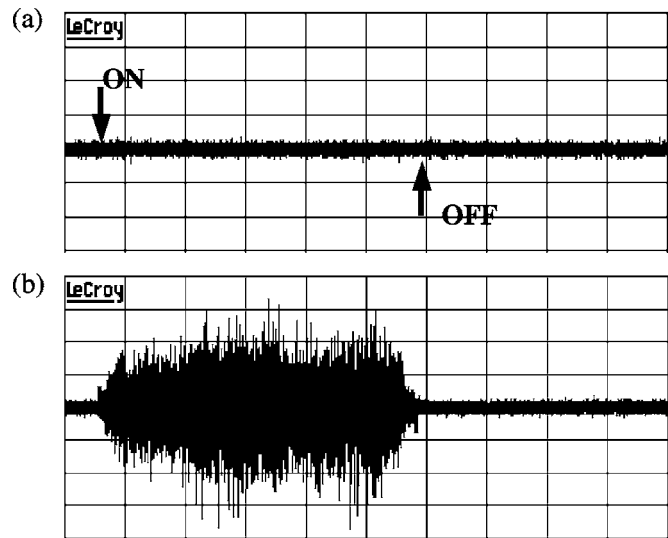
FIG. 2. Acoustic cavitation detected inside a water-filled vial prior to (a) and after (b) an injection of Optison® The start and end of the 200-kHz ultrasound pulse are shown by the arrows. The ultrasound pulse duration was 1 ms and the peak pressure amplitude was 0.2 MPa. Time/division: 0.2 ms.

jected with Optison® ($n=10$) and mice injected with saline ($n=5$) underwent the same experimental protocol.

Experiments were performed to correlate ultrasound-induced premature contractions with the extent of cavitation activity in the heart as measured with the PCD transducer. Each ultrasound exposure consisted of a 1-ms pulse of 200-kHz ultrasound delivered to the heart during diastole. Acoustic pulses were delivered during diastole because previous studies have shown that the lowest threshold for ultrasound-induced premature beats occurs when the ultrasound pulse is delivered during diastole (Dalecki *et al.*, 1993a; MacRobbie *et al.*, 1997). Peak pressure amplitudes investigated were 0.1, 0.15, 0.2, and 0.25 MPa. Each mouse received ten pulses at each pressure amplitude (i.e., a total of 40 pulses) in a randomized order. The time between each exposure pulse was ~5 s. The PCD apparatus was used to quantify the extent of cavitation activity produced in the heart by each pulse of ultrasound from the 200-kHz source. Simultaneously, an investigator blind to exposure condition observed the ECG and recorded whether or not each ultrasound pulse produced a premature contraction. At the end of each experiment, animals were euthanized under anesthesia and the heart was examined for gross hemorrhage.

III. RESULTS

A. *In vitro* results

In vitro studies primarily provided validation of the PCD technique that was later implemented with mice *in vivo*. Figure 2 shows an example of the signal received by the PCD transducer from a water vial exposed to an ultrasound pulse both prior to and after an injection of Optison®. The ultrasound pulse was at 200 kHz with pulse duration of 1 ms and pressure amplitude of 0.2 MPa. The beginning and end of the ultrasound pulse are marked by the arrows (ON and OFF, respectively). Without microbubbles present in the water vial

2960   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Rota *et al.*: Detection of cavitation in the heart
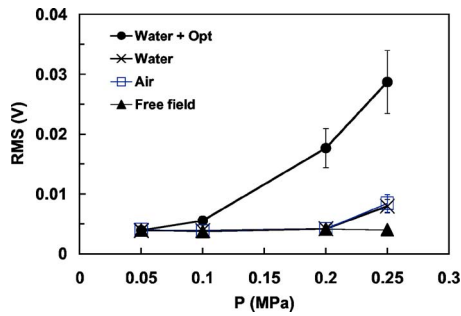
FIG. 3. Cavitation detection *in vitro*. Plotted are the mean rms (RMS) values of the output of the PCD transducer as a function of pressure amplitude of the 200-kHz ultrasound pulse. Each data point is an average of ten measurements (i.e., ten acoustic pulses) from five separate vials (total of 50 measurements). Measurements were performed with vials filled with degassed water+Optison® (solid circles), degassed water (crosses), air (open squares), or with no vial, free field measurements, (solid triangles). Error bars are standard error of the mean.

[Fig. 2(a)], no cavitation activity was detected. With microbubbles [Fig. 2(b)], significant cavitation levels were measured for the entire duration of the ultrasound pulse. Spectral analysis of the PCD output verified that the 200-kHz incident and scattered field were removed from the signal by the 2-MHz high-pass filter. The PCD output from vials with microbubbles had a broadband noise spectrum as expected from the violent collapse of bubbles in inertial cavitation.

Results of cavitation detection experiments using different vial contents are summarized in Fig. 3. This figure presents the mean of the rms (RMS) signals from the PCD transducer as a function of pressure amplitude of the incident 200-kHz ultrasound pulse. Each data point is an average of ten measurements (ten pulses/pressure amplitude) from n =5 separate vials. Thus, each data point is the mean of 50 measurements. At 0.05 MPa, RMS values were indistinguishable from the noise floor and from free field measurements (solid triangles) regardless of the vial content. This finding confirmed that the 2-MHz filter successfully removed any incident or reflected sound from the PCD output. For instance, without the 2-MHz filter, an air-filled vial (simulating the highly reflecting lung) produced a strong signal due to reflection of the 200-kHz sound (i.e., a pressure release interface). Instead, with the filter, RMS values from both air (open squares) and water (crosses) vials were negligible up to ~0.25 MPa. The small yet measurable increase in RMS found with both air and water vials at 0.25 MPa likely indicates that cavitation activity may have occurred in the water medium surrounding the vials only at this highest pressure amplitude.

In vials containing Optison® and water (solid circles), the mean of the rms (RMS) signals from the PCD was significantly increased compared to free field measurements with pressure amplitudes as low as 0.1 MPa (Fig. 3). Furthermore, RMS ms values in the presence of Optison® were substantially greater than air or water vials alone. The magnitude of the RMS of the PCD signal in the presence of Optison® increased with increasing pressure amplitude of the 200-kHz incident field.
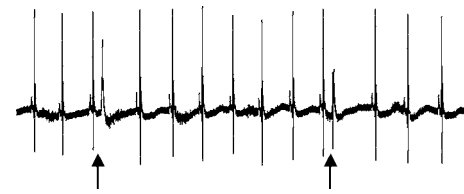
FIG. 4. ECG demonstrating ultrasound-induced premature contractions in a mouse injected with Optison® and exposed to 200 kHz ultrasound pulses. Each ultrasound pulse was of duration of 1 ms. Ultrasound pressure amplitudes were 0.15 and 0.2 MPa (first and second arrows, respectively).

B. *In vivo* results

An ECG illustrating a premature cardiac beat produced by a single ultrasound pulse is shown in Fig. 4. In the figure, two premature ventricular contractions are shown for a mouse injected with Optison® microbubbles and exposed to ultrasound. Both premature contractions developed immediately following the delivery of a 200-kHz ultrasound pulse of 1-ms duration (arrows indicate delivery of the ultrasound pulse). Pressure amplitudes were 0.15 and 0.2 MPa (first and second arrows, respectively). Typical features of a premature beat included a premature and abnormal ventricular depolarization signal (QRS), commonly followed by a compensatory pause. No evidence of gross hemorrhage to the myocardium was found in any mice.

Figure 5 depicts the mean percentage of pulses producing a premature beat as a function of pressure amplitude. For each pressure amplitude, the percentage of delivered pulses resulting in a premature cardiac contraction was determined for each mouse, and then results from all mice were used to determine the mean percentage for each exposure condition. Each data point is the mean of either 100 measurements (Optison®, $n=10$) or 50 measurements (saline, $n=5$). No premature beats were observed in any of the mice injected with saline and exposed to ultrasound (open diamonds). In comparison, in mice injected with Optison® (solid circles), the mean percentage of pulses producing a premature beat increased with increasing pressure amplitude. The increase was approximately linear up to 0.2 MPa, at which point pulse
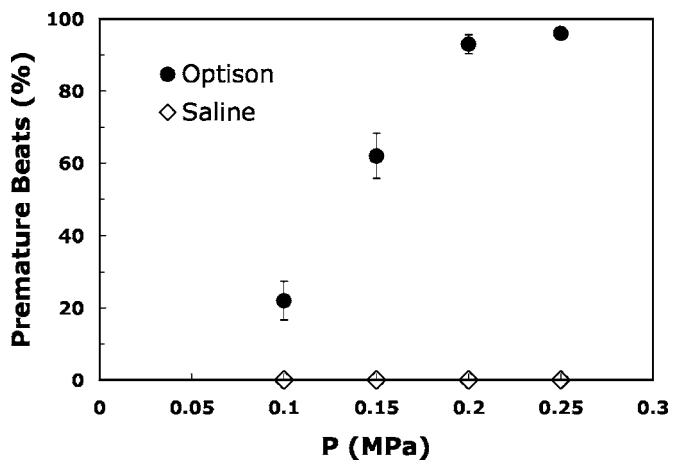
FIG. 5. Mean percentage of ultrasound pulses producing a premature beat as a function of pressure amplitude of the 200-kHz incident pulse. Pulse duration was 1 ms. Solid circles are data for mice injected with Optison® ($n=10$) and open triangles are data for mice injected with saline ($n=5$).

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Rota *et al.*: Detection of cavitation in the heart   2961
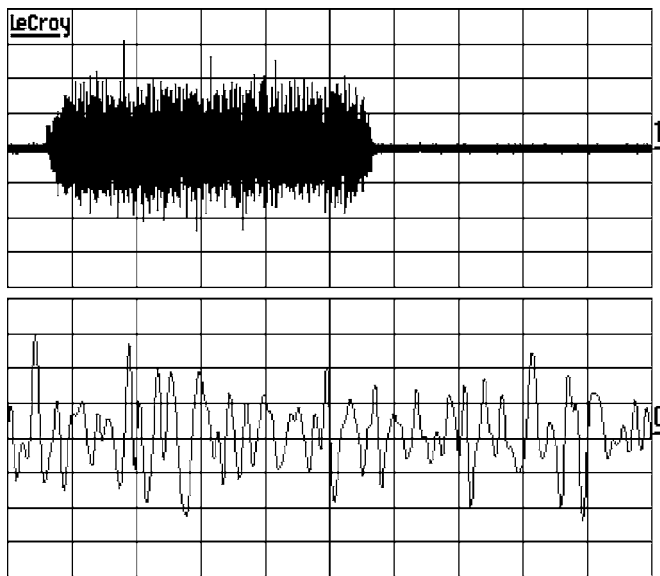
FIG. 6. Acoustic cavitation detected in the mouse heart *in vivo*. Shown is the output of the PCD transducer following delivery of an ultrasound pulse to a mouse heart with contrast agent. Ultrasound pulse was at 200 kHz with pulse duration of 1 ms and pressure amplitude of 0.2 MPa. PCD signal was filtered with a 2-MHz high-pass filter. An expanded view of the PCD signal is shown at bottom. Time/div.=0.2 ms (top).

effectiveness leveled off since the heart was nearly 100% responsive to ultrasound stimulation. Standard errors of the mean (SEM) decreased with increasing pressure amplitude indicating less variance in cardiac response at the highest acoustic pressures.

During the ultrasound exposures, the PCD transducer was used to quantify the extent of cavitation activity in the heart. An example of cavitation detected in vivo is shown in Fig. 6. In this example, the mouse was injected with Optison[®] microbubbles and exposed to a pulse of ultrasound (200 kHz, 1 ms) at a pressure amplitude of 0.2 MPa. Figure 7 depicts the results of cavitation measurements *in vivo* for
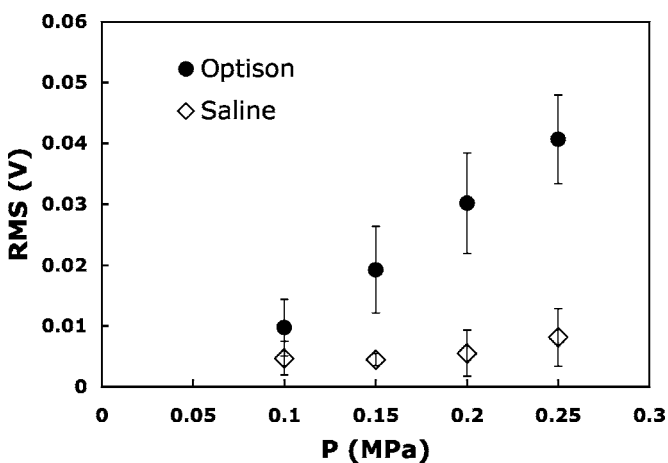
FIG. 7. Cavitation detection *in vivo*. Plotted are the mean rms (RMS) values of the output of the PCD transducer as a function of pressure amplitude of the 200-kHz ultrasound pulse. Pulse duration was 1 ms. Results are shown for both mice injected with Optison[®] (solid circles) and mice injected with saline (open diamonds). Each data point is the mean of either 100 measurements (Optison[®], $n=10$) or 50 measurements (Saline, $n=5$). Error bars are standard error of the mean.
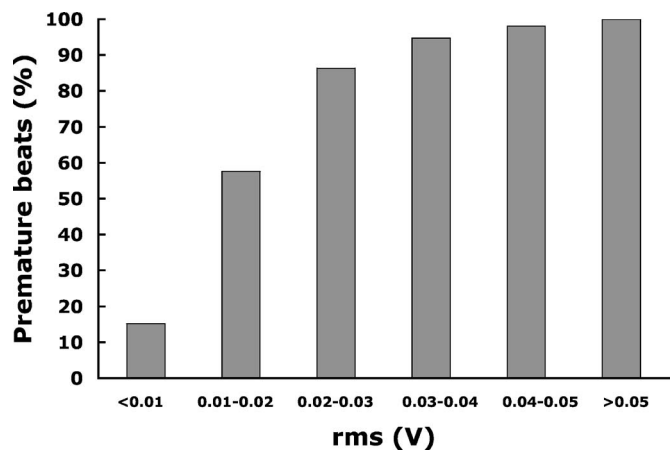
FIG. 8. Correlation between the amplitude of the rms of PCD output and premature beats in mice injected with Optison[®] ($n=10$). Ultrasound frequency and pulse duration were 200 kHz and 1 ms, respectively. Pressure amplitudes were 0.1, 0.15, 0.2, and 0.25 MPa.

mice injected with either saline (open diamonds) or Optison[®] (solid circles). In the figure, the mean (RMS) of the PCD transducer output is plotted as a function of pressure amplitude of the incident 200-kHz ultrasound pulse. Each data point is the mean of either 100 measurements (Optison[®], number of mice=$n=10$) or 50 measurements (saline, $n=5$). In mice injected with Optison[®], the RMS of the PCD signal increased linearly with increasing pressure amplitude of the 200-kHz incident ultrasound pulse ($R^2=0.999$). In animals injected with saline, RMS measurements were indistinguishable from the noise floor. As in experiments *in vitro*, there was a slight increase in the RMS at 0.25 MPa that presumably is indicative of cavitation in the water only at this highest pressure amplitude.

The correlation between amplitude of the rms PCD signal and ultrasound-induced premature beats in mice injected with Optison[®] is shown in Fig. 8. In the figure, the mean percentage of pulses producing a premature beat is plotted as a function of the rms of the PCD output for five ranges of rms values. The data were pooled from all animals injected with Optison[®] ($n=10$). When small rms values were detected (<0.01 V), few pulses resulted in a premature beat (15.1%). However, an increase in the amplitude of the PCD output correlated with a monotonic increase in the percentage of pulses producing a premature beat. Thus, the figure indicates that as the extent of cavitation activity increased (as quantified by the rms of the PCD signal), the likelihood of producing a premature cardiac contraction increased.

IV. DISCUSSION

Microbubbles undergoing acoustic cavitation act as secondary sources of sound. Detection and analysis of these acoustic emissions is often used as a way to monitor the occurrence of acoustic cavitation and distinguish between noninertial and inertial cavitation (Neppiras, 1980; Apfel, 1981). At low exposure pressures, the spectrum of the acoustic emissions from nonlinear oscillations of microbubbles includes harmonics, ultraharmonics, and subharmonics of the exposure frequency (Apfel, 1981; Leighton, 1994). Spectra

of acoustic emissions for inertial cavitation activity are characterized by a broadband distribution of frequencies, white noise, attributable to the violent collapse of bubbles (Akulichev, 1971; Neppiras, 1980; Apfel, 1981; Ilychev *et al.*, 1989; Leighton, 1994; Everbach *et al.*, 1997).

Passive cavitation detection has been employed in numerous studies to detect and monitor acoustic cavitation *in vitro* (e.g., Atchley *et al.*, 1988; Roy *et al.*, 1990; Holland and Apfel, 1990; Crum *et al.*, 1992; Everbach *et al.*, 1997). A limited number of studies have used the PCD technique to detect cavitation *in vivo*. Investigations have used the technique to demonstrate the production of acoustic cavitation in tissues exposed to lithotripter fields (e.g., Bailey *et al.*, 2005; Zhong *et al.*, 1997) or high intensity focused ultrasound (e.g., Thomas *et al.*, 2005). Porter *et al.*, (2001) used a PCD technique to detect intramyocardial cavitation in open-chest dogs exposed to pulsed ultrasound at 1.7 MHz after intravenous injections of a contrast agent.

In this investigation, we used a PCD technique to test directly the role of cavitation in ultrasound-induced premature cardiac contractions in mice injected with microbubble contrast agents. The design of this study provided an ideal opportunity to demonstrate a direct correlation of an ultrasound bioeffect with cavitation *in vivo*. A premature contraction is an "all-or-nothing" bioeffect that can be observed noninvasively immediately after delivery of an ultrasound pulse. Thus, for each acoustic pulse delivered to the heart, we were able to immediately correlate the extent of cavitation activity measured by the PCD transducer with the effectiveness of the pulse in producing a premature cardiac contraction.

Initial measurements of acoustic cavitation detection were performed *in vitro* to validate the measurement technique and provide insight necessary for interpreting the measurements of cavitation activity in mice *in vivo*. *In vitro* tests provided evidence that neither the incident nor the reflected sound fields were detected in the high-pass filtered (2 MHz cutoff) output signal of the PCD transducer. The filtered PCD signal amplitudes measured without any vial in the field and with either air or water vials were negligible up to ~0.2 MPa (Fig. 3). In comparison, in vials containing Optison®, the magnitude of the mean rms of the PCD signal increased with increasing pressure amplitude of the 200-kHz incident field (Fig. 3).

No premature contractions were observed in any mice injected with saline and exposed to ultrasound (Fig. 5). Also, in mice injected with saline, there was no significant increase in the RMS of the PCD signal with increasing exposure amplitude (Fig. 7). Thus, in mice injected with saline and exposed to ultrasound, no cavitation activity was detected and no premature contractions were produced. In comparison, ultrasound was an effective stimulus of premature beats in mice injected with Optison® (Fig. 5). For mice injected with contrast agent, both the mean percentage of ultrasound pulses producing a premature beat (Fig. 5) and the RMS of the PCD signal (Fig. 7) increased with increasing pressure amplitude. Furthermore, there was a direct correlation between the extent of cavitation, quantified by the PCD transducer signal, and the likelihood of producing a premature contraction with ultrasound (Fig. 8).

In this investigation, a low-acoustic frequency, long pulse duration (200 kHz, 1 ms), and low-contrast agent dose (3 μL/bolus) were chosen to minimize gross vascular damage to the heart while still maintaining the effectiveness of the ultrasound in producing premature contractions (Rota, 2005). Indeed, there was no evidence of gross hemorrhage on the surface of the heart in any of the mice injected with either saline or contrast agents and exposed to ultrasound pulses (200 kHz, 1 ms) with pressure amplitudes ranging from 0.1 to 0.25 MPa. This is in contrast to recent reports of myocardial microvessel rupture in mice and rats injected with contrast agents and exposed to ultrasound at diagnostic frequencies (Li *et al.*, 2003; 2004; Dalecki *et al.*, 2005b; Miller *et al.*, 2005). Li *et al.* (2003, 2004) determined a pressure threshold of 0.5 MPa for production of microvascular damage in rats injected with microbubble contrast agent and exposed to a clinical diagnostic device operating at 1.7 MHz. In rats injected with contrast agent, Miller *et al.* (2005) did not find a direct correlation between ultrasound-induced myocyte injury and the production of premature beats. The ability of ultrasound at 200 kHz in the presence of microbubble contrast agents to stimulate premature beats without any gross evidence of vascular damage to the myocardium suggests that ultrasound-induced premature contractions are a highly sensitive indicator of cavitation *in vivo*.

V. CONCLUSION

This study implemented a passive cavitation detection technique to measure acoustic cavitation in the murine heart *in vivo* when microbubble contrast agents were present in the blood. There was a direct correlation between the amplitude of the passive cavitation detector output and the likelihood of producing a premature beat with a pulse of ultrasound. These results provide direct evidence in support of inertial cavitation as a mechanism for ultrasound-induced premature cardiac contractions in the presence of microbubble contrast agents.

ACKNOWLEDGMENTS

AIUM (American Institute of Ultrasound in Medicine) (**2000**). "Mechanical bioeffects from diagnostic ultrasound: AIUM consensus statements," J. Ultrasound Med. **19**, 120–142.

Akulichev, V. A. (**1971**). "Pulsations of cavitation voids," *High Intensity Ultrasonic Fields*, edited by L. D. Rozenberg (Plenum Press, New York).

Apfel, R. E. (**1981**). "Acoustic cavitation," Methods Exp. Phys. **18**, 355–412.

Atchley, A. A., Frizzell, L. A., Apfel, R. E., Holland, C. K., Madenshetty,

J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Rota *et al.*: Detection of cavitation in the heart   2963

S., and Roy, R. A., (**1988**). "Thresholds for cavitation produced in water by pulsed ultrasound," Ultrasonics **26**, 280–285.

Bailey, M. R., Pishchalnikov, Y. A., Sapozsnikov, O. A., Cleveland, R. O., McAteer, J. A., Miller, N. A., Pishchalnikova, I. R., Connors, B. A., Crum, L. A., and Evan, A. P. (**2005**). "Cavitation detection during shock-wave lithotripsy," Ultrasound Med. Biol. **31**, 1245–1256.

Crum, L. A., Roy, R. A., Dinno, M. A., Church, C. C., Apfel, R. E., Holland, C. K., and Madenshetty, S. I. (**1992**). "Acoustic cavitation produced by microsecond pulses of ultrasound: A discussion of some selected results," J. Acoust. Soc. Am. **91**(2), 1113–1119.

Dalecki, D. (**2005a**). "Biological effects of microbubble ultrasound contrast agents," in *Medical Radiology-Diagnostic Imaging. Volume: Contrast Media in Ultrasonography, Basic Principles and Clinical Applications*, edited by Emilio Quaia (Springer-Verlag, Berlin), pp. 77–85.

Dalecki, D., Keller, B. B., Carstensen, E. L., Neel, D. S., Palladino, J. L., and Noordergraaf, A. (**1991**). "Thresholds for premature ventricular contractions in frog hearts exposed to lithotripter fields," Ultrasound Med. Biol. **17**, 341–346.

Dalecki, D., Keller, B. B., Raeman, C. H., and Carstensen, E. L. (**1993a**). "Effects of pulsed ultrasound on the frog heart: I. Thresholds for changes in cardiac rhythm and aortic pressure," Ultrasound Med. Biol. **19**, 385–390.

Dalecki, D., Raeman, C. H., and Carstensen, E. L. (**1993b**). "Effects of pulsed ultrasound on the frog heart: II. An investigation of heating as a potential mechanism," Ultrasound Med. Biol. **19**, 391–398.

Dalecki, D., Raeman, C. H., Child, S. Z., and Carstensen, E. L. (**1997a**). "Effects of pulsed ultrasound on the frog heart: III. The radiation force mechanism," Ultrasound Med. Biol. **23**, 275–285.

Dalecki, D., Raeman, C. H., Child, S. Z., Cox, C., Francis, C. W., Meltzer, R. S., and Carstensen, E. L. (**1997b**). "Hemolysis *in vivo* from exposure to pulsed ultrasound," Ultrasound Med. Biol. **23**, 307–313.

Dalecki, D., Rota, C., Raeman, C. H., and Child, S. Z. (**2005b**). "Premature cardiac contractions produced by ultrasound and microbubble contrast agents in mice," ARLO **6**, 221–225.

Delius, M., Hoffmann, G., Steinbeck, G., and Conzen, P. (**1994**). "Biological effects of shock waves: Induction of arrhythmia in piglet hearts," Ultrasound Med. Biol. **20**, 279–285.

Everbach, E. C., Makin, I. R. S., Azadniv, M., and Meltzer, R. S. (**1997**). "Correlation of ultrasound-induced hemolysis with cavitation detection *in vitro*," Ultrasound Med. Biol. **23**(4), 619–624.

Holland, C. K., and Apfel, R. E. (**1990**). "Thresholds for transient cavitation produced by pulsed ultrasound in a controlled nuclei environment," J. Acoust. Soc. Am. **88**, 2059–2069.

Ilyichev, V. I., Koretz, V. L., and Melnikov, N. P. (**1989**). "Spectral charac-

teristics of acoustic cavitation," Ultrasonics **27**, 357–361.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic, New York).

Li, P., Armstrong, W. F., and Miller, D. L. (**2004**). "Impact of myocardial contrast echocardiography on vascular permeability: comparison of three different contrast agents," Ultrasound Med. Biol. **30**(1), 83–91.

Li, P., Cao, L., Dou, C., Armstrong, W. F., and Miller, D. L. (**2003**). "Impact of myocardial contrast echocardiography on vascular permeability: An *in vivo* dose response study of delivery mode, pressure amplitude and contrast dose," Ultrasound Med. Biol. **29**(9), 1341–1349.

MacRobbie, A. G., Raeman, C. H., Child, S. Z., and Dalecki, D. (**1997**). "Thresholds for premature contractions in mouse hearts exposed to pulsed ultrasound," Ultrasound Med. Biol. **23**, 761–765.

Miller, D. L., Peng, L., Dou, C., Gordon, D., Edwards, C. A., and Armstrong, W. F. (**2005**). "Influence of contrast agent dose and ultrasound exposure on cardiomyocyte injury induced by myocardial contrast echocardiography in rats," Radiology **237**, 137–143.

NCRP (National Council on Radiation Protection and Measurements) (**2002**). "Exposure Criteria for Medical Diagnostic Ultrasound. II. Criteria Based on All Known Mechanisms," NCRP, Bethesda, MD.

Neppiras, E. A. (**1980**). "Acoustic cavitation," Phys. Rep. **61**(3), 159–251.

Porter, T. R., Everbach, C., Kricsfeld, D., and Xie, F. (**2001**). "Myocardial cavitational activity during continuous infusion and bolus intravenous injections of perfluorocarbon-containing microbubbles," J. Am. Soc. Echocardiogr **14**, 618–625.

Rota, C. (**2005**). "Cardiac arrhythmias produced by ultrasound and contrast agents," Ph.D. thesis, University of Rochester, Rochester, NY.

Roy, R. A., Madanshetty, S. I., and Apfel, R. E. (**1990**). "An acoustic backscattering technique for the detection of transient cavitation produced by microsecond pulses of ultrasound," J. Acoust. Soc. Am. **87**(6), 2451–2458.

Smailys, A., Dulevicius, Z., Muckus, K., and Dauksa, K. (**1981**). "Investigation of the possibilities of cardiac defibrillation by ultrasound," Resuscitation **9**, 233–242.

Thomas, C. R., Farney, C. H., Coussios, C. C., Roy, R. A., and Holt, R. G. (**2005**). "Dynamics and control of cavitation during high-intensity focused ultrasound application," ARLO **6**, 182–187.

van der Wouw, P. A., Brauns, A. C., Bailey, S. E., Powers, J. E., and Wilde, A. A. (**2000**). "Premature ventricular contractions during triggered imaging with ultrasound contrast," J. Am. Soc. Echocardiogr **13**, 288–294.

Young, F. (**1989**). *Cavitation* (McGraw-Hill, New York).

Zhong, P., Cioanta, I., Cocks, F. H., and Preminger, G. M. (**1997**). "Inertial cavitation and associated acoustic emission produced during electrohydraulic shock wave lithotripsy," J. Acoust. Soc. Am. **101**, 2940–2950.

2964   J. Acoust. Soc. Am., Vol. 120, No. 5, November 2006

Rota *et al.*: Detection of cavitation in the heart

Advantages of the Hilbert Huang transform for marine mammals signals analysis

Olivier Adam[a]

Laboratoire Images, Signaux et Systèmes Intelligents groupe Ingénierie des Signaux Neuro-Sensoriels
Université Paris 12 – 61 av de Gaulle, 94000 Creteil, France

While marine mammals emit variant signals (in time and frequency), the Fourier spectrogram appears to be the most widely used spectral estimator. In certain cases, this approach is suboptimal, particularly for odontocete click analysis and when the signal-to-noise ratio varies during the continuous recordings. We introduce the Hilbert Huang transform (HHT) as an efficient means for analysis of bioacoustical signals. To evaluate this method, we compare results obtained from three time-frequency representations: the Fourier spectrogram, the wavelet transform, and the Hilbert Huang transform. The results show that HHT is a viable alternative to the wavelet transform. The chosen examples illustrate certain advantages. (1) This method requires the calculation of the Hilbert transform; the time-frequency resolution is not restricted by the uncertainty principle; the frequency resolution is finer than with the Fourier spectrogram. (2) The original signal decomposition into successive modes is complete. If we were to multiply some of these modes, this would contribute to attenuate the presence of noise in the original signal and to being able to select pertinent information. (3) Frequency evolution for each mode can be analyzed as one-dimensional (1D) signal. We not need a complex 2D post-treatment as is usually required for feature extraction. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2354003]

I. INTRODUCTION

This paper describes an endeavor to use the Hilbert Huang transform (HHT) to analyze bioacoustic signals, particularly vocalizations and clicks emitted by marine mammals.

Our aim is not to generalize the results of this approach when applied to the emitted signals of all marine mammals but, rather, to show its efficiency in analyzing certain vocalizations and clicks. The objective is to contribute to extract some features from these signals, as present techniques used for both detecting signals emitted by marine mammals and for defining an acoustic signature call for improvements (Erbe, 2000; Tiemann and Porter, 2003).

Using passive acoustics, the final applications could involve the tracking of certain "deep water" species (such as sperm whales "physeter macrocephalus"), supplementing visual observations, reducing significantly the percentage of missed trackline groups (Barlow and Shannon, 2004), and estimating the number of animals in a specific marine zone.

Bioacoustic signals emitted by whales present a variability of characteristics between species and between different individuals within the same species (Nakamura and Akamatsu, 2004). Moreover, these signals could be complex when numerous individuals emit sounds simultaneously. Also, these signals deteriorate with the presence of noise (natural or anthropogenic); this signal-to-noise ratio is liable to vary during the entire recording. The easiest approach for detecting and extracting signals from continuous recordings is to use an energy function. This detector is not optimal

when certain features of these signals are known *a priori*. To improve its performance, different signal processing algorithms could be proposed. Some of them provide a time-frequency or time-scale representation. These algorithms are broken down into three steps: step (1) signal preprocessing; step (2) selection of a decomposition method; and step (3) feature extraction. (1) Step 1 focuses on pertinent information present in the recordings. Usually, it consists of selecting the frequency bandwidth corresponding to signals emitted by the considered species. Each marine mammal species emits signals with specific frequency range; this bandwidth is considered for analysis. Signals must also be filtered for noise, reverberation, and multipath effects. (2) Different methods can be used during step 2: the Fourier spectrogram, the AR model, the wavelet transform, etc. (3) The interpretation of the time-frequency or time-scale representation obtained from step 2 could be achieved manually or inserted in an automatic decision algorithm: statistical approach, artificial neural networks, hidden Markov models, etc.

This paper focuses on the second step. This step could be particularly important in this process because it has consequences on the third step. We will evaluate the advantages of the HHT by considering the consequences on step 3.

Time-frequency or time-scale representation is an integral step in the segmentation and the characterization of the signals. Different methods are evaluated but often the Fourier spectrogram is the final choice for analyzing marine mammal recordings. Recently, Mellinger (2005) used this method to detect different species. Jarvis *et al.* (2003) also proposed the Fourier spectrogram for detecting different species of toothed whales. Because this approach is limited by the Heisenberg uncertainty principle, the authors have to

a)Electronic mail: adam@univ-paris12.fr

make choices based on their applications. They set a noise variable threshold (NVT) that adapts to the background noise. For example, in (Jarvis *et al.*, 2003), the resolutions in time and frequency are chosen, respectively, at about 10 ms and 50 Hz. Next, Fourier spectrograms are compared to a threshold and a binary time-frequency representation is provided—this representation will be used in step 3.

Because this threshold is often set manually in the beginning of the process and kept constant for the complete analysis of all the recordings, it seems difficult to characterize a complete marine mammal vocalization or click. Indeed, in this case, continuous recordings present nonstationary characteristics: the signals emitted by marine mammals are intermittent and when these signals are present, their features vary in time. Moreover, the signal-to-noise ratio varies during the recordings. In the end, this threshold could always be suboptimal.

However, some local information could be extracted from Fourier time-frequency representations (Altes, 1980; Kadambe and Orr, 1995; Kuperman *et al.*, 2001; Hory *et al.*, 2002). For vocalizations, Matthews *et al.* (1999) extracted pertinent characteristics such as the start, end (duration), minimum, maximum, and central frequencies, and number of inflections. Gillespie (2004) provides species classification from Fourier spectrogram results. Oswald *et al.* (2003) proposed using multivariate discriminating function analysis for classifying vocalizations of different whales. This classification is also based on the results of the Fourier spectrogram. Backpropagation neural networks could be used for detecting features from the whales' spectrograms, even when obstructed by noise (Erbe, 2000; Potter *et al.*, 1994).

For signals emitted by marine mammals, it appears that the Fourier spectrogram is the most known and applied approach for step 2, i.e., after the filtering of the signals and before feature extraction.

Note that the Fourier approach is adapted for harmonic signals. Sinusoids are sufficient for analyzing vocalizations from gray whales (Dalheim, 1984) but not for odontocete clicks.

Cohen (1995) has shown that the Fourier estimator is limited to linear time-frequency analysis. In many cases, vocalizations are analyzed despite these restrictions, such as in human speech—vowels could be considered quasistationary ($\sim$20 ms) for voice phonemes. These hypotheses are too strong for marine mammal signals due to the presence of brief vocalizations or transient clicks (Ioana and Quinquis, 2004). Thus, resulting spectrograms have to be interpreted using precaution. In the end, the Fourier estimator is maybe not the optimal estimator for the analysis of continuous recordings when the goal is to extract certain features from the time-frequency representation of odontocete clicks; and moreover, when noise is susceptible to variant properties.

Different signal processing approaches are proposed here and considered more appropriate than Fourier estimators for nonstationary signals analysis (Boashash, 2003). These methods consist of decomposing the original signals into a specific vector basis, similar to the use of Fourier estimators (decomposition into complex exponentials). The wavelet transform is one example (Meyer, 1993). Thus, the

first step for providing these time-frequency or time-scale representations is to determine what kind of vector basis is necessary and optimal for decomposition. For different studies, the vector basis is chosen according to the selected application: The same basis cannot be used in analyzing signals from different marine mammals or different individuals of the same species. Various authors have illustrated the advantages of the wavelet family used for their own specific applications—Fargues and Bennett (1995) justified the use of the Morlet mother wavelet; in 1998, Huele and De Haes preferred the analysis using the packet wavelet; and Lopatka *et al.* (2005) presented better results obtained with the Daubechies mother wavelet.

Introducing the HHT proposed by Huang *et al.* (1998) appears to be an alternative for characterizing these bioacoustic signals. This method does not need any vectors for decomposition. Signals are decomposed into themselves; this means that no specific vectors are required and, consequently, no bias is introduced during the decomposition process. Theoretical details are briefly presented in the next section. Afterward, the results are presented on simulated signals and real marine mammal recordings. The conclusion highlights the significance of our results.

II. METHOD

The Hilbert Huang transform was introduced by Huang *et al.* (1998). This method can be presented in two phases. First, signals are decomposed into a finite set of intrinsic mode functions (IMF). This first step is called empirical mode decomposition (EMD). Then, the Hilbert transform (HT) is applied and a time-frequency representation is performed.

A. Empirical mode decomposition

Decomposition consists of extracting monocomponent contributions, called IMF, from the signals based on the localization of their extrema. Upper and lower envelopes are calculated; then the average of these two envelopes is calculated. A residual signal is obtained from the difference between the original signal and its mean. This process, called the *sifting process*, is repeated until the final mean satisfies two conditions: (1) the number of extrema and the number of zero crossings must either equal zero or differ at most by one, and (2) at any point, the mean value of the two envelopes is equal to zero (Huang, 1998). The signal s is decomposed by

$$s[n] = \sum_{i=1}^{M} c_i[n] + T_M[n], \tag{1}$$

with c_i the ith mode of the signal. c is also called the intrinsic mode function (IMF). T_M is the residue.

Note that this decomposition is complete. It means that the sum of all modes c_i equals the original signal. The selection of some of the modes corresponds to filtering the signal (Flandrin *et al.*, 2003).

Olivier Adam: Hilbert Huang transform bioacoustic signal analysis

The number of modes depends on the shift time-window length. But contrary to the Fourier transform, the frequency resolution does not depend on the time-window width.

B. Hilbert transform

A time-frequency representation can be obtained from the modes. Huang *et al.* (1998) proposed the use of the Hilbert transform

$$s_H[n] = \text{HT}(s[n]) = s[n] \otimes \frac{1}{\pi n}. \qquad (2)$$

Let us define the analytic signal

$$s_A[n] = s[n] + js_H[n] \qquad (3)$$

with $j^2 = -1$.

s_A could be written

$$s_A[n] = a[n]e^{j\theta[n]} \qquad (4)$$

with magnitude $a[n] = \sqrt{s^2[n] + s_H^2[n]}$ and phase $\theta[n] = \arctan s_H[n]/s[n]$.

From the Hilbert transform, the frequency is defined by Eq. (5)

$$f[n] = \frac{1}{2\pi} \frac{d\theta[n]}{dn}. \qquad (5)$$

This definition of the frequency, called instantaneous frequency, prevents limitations of the uncertainty principle. The representation of the triplet $\{n, f_i[n], a_i[n]\}$ in the time-frequency plane is provided for each IMF $(i = 1, \ldots, M)$. The Hilbert spectrum $H[n, f]$ is expressed as the sum of the Hilbert spectrums of each IMF.

The margin spectrum is defined at time $n = l/F_s$

$$h\left(\frac{l}{F_s}\right) = \sum_{f=f_0}^{f_1} H\left[\frac{l}{F_s}, f\right] \qquad (6)$$

with f_0 and f_1 the considered frequency range. F_s is the sampling frequency.

III. RESULTS

Before showing results on real recordings, we illustrate the performance of this method on simulated signals.

A. About the time-frequency resolution

Let us define the following chirp with a linear frequency varying from $0.01F_s$ to $0.3F_s$, where F_s is the sampling frequency. Figure 1 shows the comparison of three time-frequency representations.

First, the Fourier spectrogram has been calculated. Note that the rectangular or Bartlett window could be applied on the original signal before performing the Fourier transform because we need only focus on the maximum of the main frequency. The Hanning and Hamming windows are not justified in our case because their use would minimize the secondary lobes effects and not focus on the maximum position. Figure 1(a) shows results obtained with a Bartlett time window with the length $256/F_s$; the overlapping is 97%. The frequency resolution is around $4 \cdot 10^{-3} F_s$.
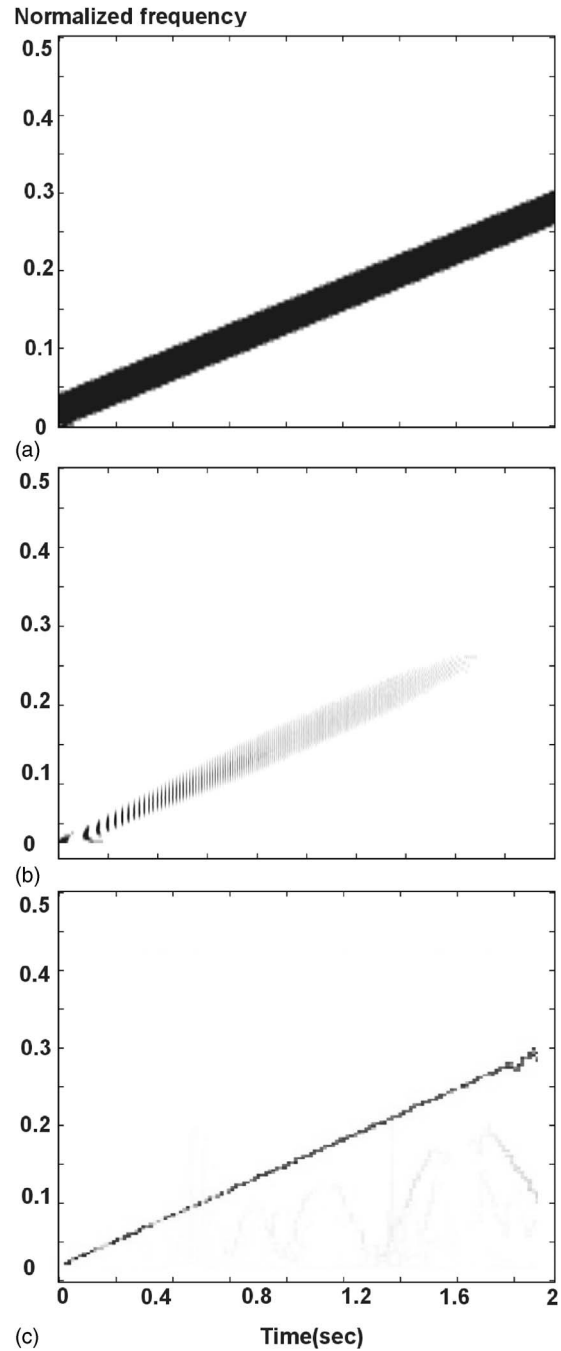
FIG. 1. Time—Frequency representation of a chirp. (a) Fourier spectrogram (time-window = 0,256 sec; overlapping = 97%). Illustration of the Heisenberg uncertainty principle. (b) Morlet wavelet transform. (c) Hilbert transform.

Second, Fig. 1(b) represents results obtained with the wavelet transform. Different wavelet families were tested—Mexican hat, Morlet, Haar, Daubechies—and finally our choice, the Morlet mother wavelet because it is well adapted for the analysis of this chirp (signal with a large frequency variation).

Note that the Morlet wavelet, defined in (Goupillaud *et al.*, 1984) is

$$\psi(t) = e^{-j2\pi f_0 t} e^{-(t/\sigma)2} \qquad (6')$$

with a the modulation term and σ the scale parameter which is used for fixing the window length.
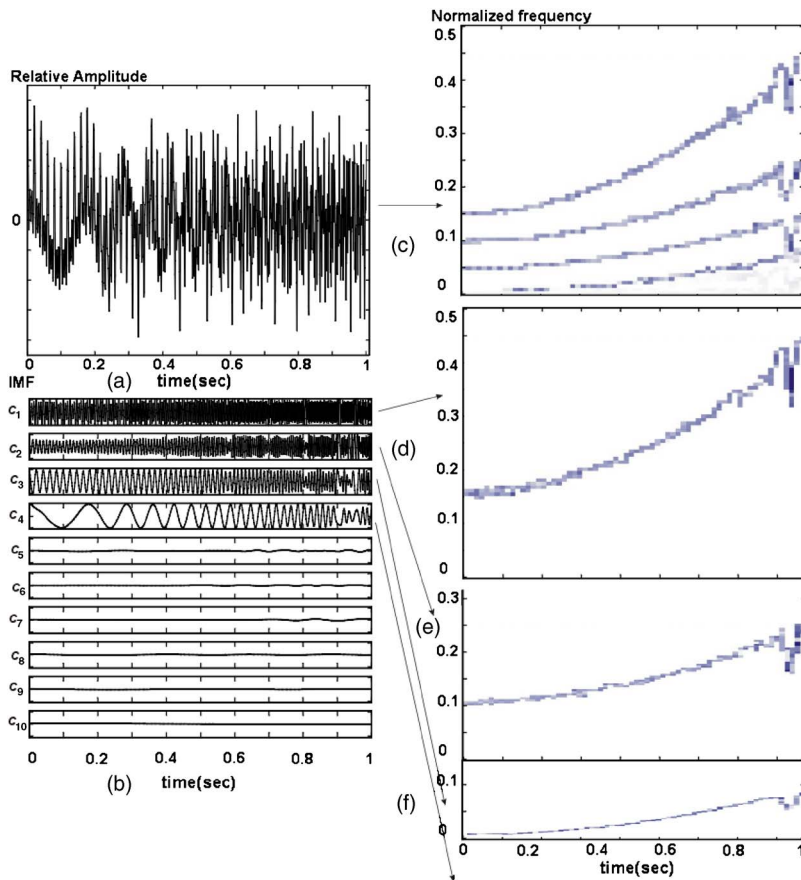
FIG. 2. (Color online) Illustration of the empirical mode decomposition of four chirps and the time-frequency representation obtained from the Hilbert transform. (a) Original signal. (b) EMD generates ten modes. (c) Time-frequency representation of the original signal. (d–f) Time-frequency representation of the first, second, third…chirp.

As defined for the HHT, we can calculate the energy of the spectrum at a given time $n = l/F_s$

$$E\left(\frac{l}{F_s}\right) = \sum_{f=f_0}^{f_1} c_w^2\left[\frac{l}{F_s}, f\right] \quad (7)$$

with c_w the wavelet coefficients. The energy E is calculated from the coefficients corresponding to the frequency band-width $[f_0, f_1]$.

Third, the last representation shows the instantaneous frequency calculated from the Hilbert transform. The HHT gives a finer time-frequency resolution than the two other methods, as we can see in Fig. 1(c).

This is clearly one of the advantages of the HHT.

B. Application on a combination of different chirps

Different kinds of chirps are used in many underwater acoustics applications: geophysics, localization, etc. These signals have finite length (usually short time) and present a linear variation of their frequencies. Some applications need chirps whose slope of frequency curves are convex or concave. These specific signals could be used for the definition of the model of natural sounds (Lewicki, 2002).

In this section, the goal is not to compare the results obtained with the HHT with other time-frequency representations, as shown previously. Rather, we wish to illustrate the first HHT step: the empirical mode decomposition.

Thus, we create a simulated signal which is the sum of four different chirps (the frequency of these signals increases in time). This signal is shown in Fig. 2(a). Ten IMF are

calculated from the original signal. The first four IMF contain the pertinent information from the four chirps. The following IMF are present due to side effects and in part because of the presence of noise. We note that the IMF are ordered from the high-frequency chirp to the lowest frequency chirp as shown in Fig. 2(b).

Then, the time-frequency representations are performed. Figure 2(c) shows results from the original signal. We clearly distinguish the presence of the four chirps. Figures 2(d)–2(f) are, respectively, the Hilbert transform of the first IMF, second IMF, third IMF (we did not add the fourth IMF to simplify the illustration). Because the IMF are mono-component signals, each time-frequency representation corresponds to the frequency variation of one chirp.

This result is another advantage of the HHT.

C. Application on signals emitted by whales

After testing the HHT on simulated signals, we have evaluated the performance on real signals. We would like to apply this method to whale signals analysis. Like in speech processing, we can distinguish two classes of signals: vocalization and clicks. The first class contains all harmonic signals. This is the case of vocalizations of killer whales, humpback whales, gray whales, etc. The second class contains only transient pulses, such as clicks. These clicks are repetitive, like those emitted by killer whales, dolphins, belugas, sperm whales, etc.

1. Illustration of killer whale vocalizations

Some vocalizations emitted by whales are similar to simulated chirps (Lewicki, 2002). These chirps have variant shapes; meaning that the vocalizations could be approximated with a linear or concave varying frequency. The duration fluctuates from 10 ms (killer whales) to more than 5 s (humpback whales).

In this part of our work, we have considered recordings of signals emitted by a group of killer whales. We isolated one specific vocalization to illustrate the performance of the HHT. Figure 3 shows the comparison between the three time-frequency representations.

Even if the resolution both in time and frequency is finer with the Hilbert spectrum, the presence of noise in the recordings perturbs the calculation of the instantaneous frequency, as we can see in Fig. 3(c). The presence of noise degrades the Fourier spectrogram as well [Fig. 3(a)]. The wavelet transform appears more robust [Fig. 3(b)].

To improve the time-frequency representation provided by the HHT, we have to focus on the IMF decomposition. In this particular example, the original vocalization containing noise is decomposed into 14 IMF [Fig. 4(a)]. In the last IMF, the noise is diffused. Thus, taking into account only the first IMF is equivalent to filtering the original vocalization.

Figure 4(c) shows the time frequency of the part of the vocalization calculated from the first mode. This signal is without noise. Feature extraction from this time-frequency representation will be easier; this step is illustrated by characterizing the vocalization fluctuations. We have chosen to fit the successive values of the instantaneous frequency by a polynomial regression [Fig. 4(d)].

This result is yet another advantage of the HHT.

2. Illustration of sperm whales creaks

As mentioned in the Introduction, the Fourier estimators are particularly suited to harmonic signals. Results could be difficult to interpret when applied to transient signals, as is the case with clicks emitted by sperm whales or beaked whales. Detection of clicks in continuous recordings is still a challenge for detection algorithms. The non-trivial tasks are threefold: (1) to distinguish the presence of clicks, providing the best time-frequency resolution possible is generally the goal of methods based on time-frequency representations, especially for time-delay measurements (Kuperman *et al.* 2001; Jarvis *et al.* 2003); (2) to identify the successive pulses in the click and separate the complete click from its echoes, and (3) to extract pertinent patterns.

Different detectors have been proposed as alternatives to those based on the Fourier estimators. The wavelet transform seems more suitable for the decomposition of these kinds of signals.

In this section, we illustrate the advantages of the HHT when compared to the Fourier spectrogram and the wavelet transform for analysis of these signals and particularly of sperm whale clicks.

Sperm whales emit different kinds of clicks: regular clicks, creaks, codas (Wahlberg, 2002). Because regular clicks have high directionality, source levels up to ≥ 200 dB
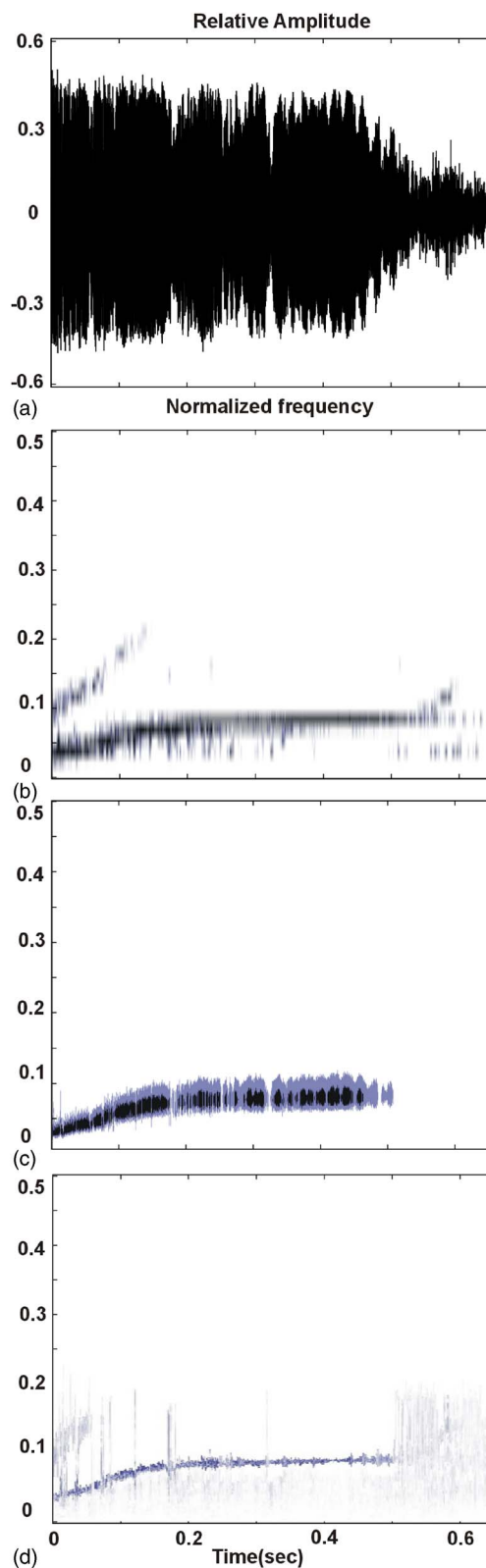
FIG. 3. (Color online) Time-frequency representation of a killer whale vocalization. (a) Time signal. (b) Performed with the Fourier transform. (c) Performed with the Morlet wavelet transform. (d) Performed with the HHT.

re: 1 μPa and are repetitive (around one per second), the performance of each technique is evaluated when applied to regular clicks. It is more difficult to find results on creaks detection in existing specialized literature, in spite of the importance of these signals. Creaks contain successive
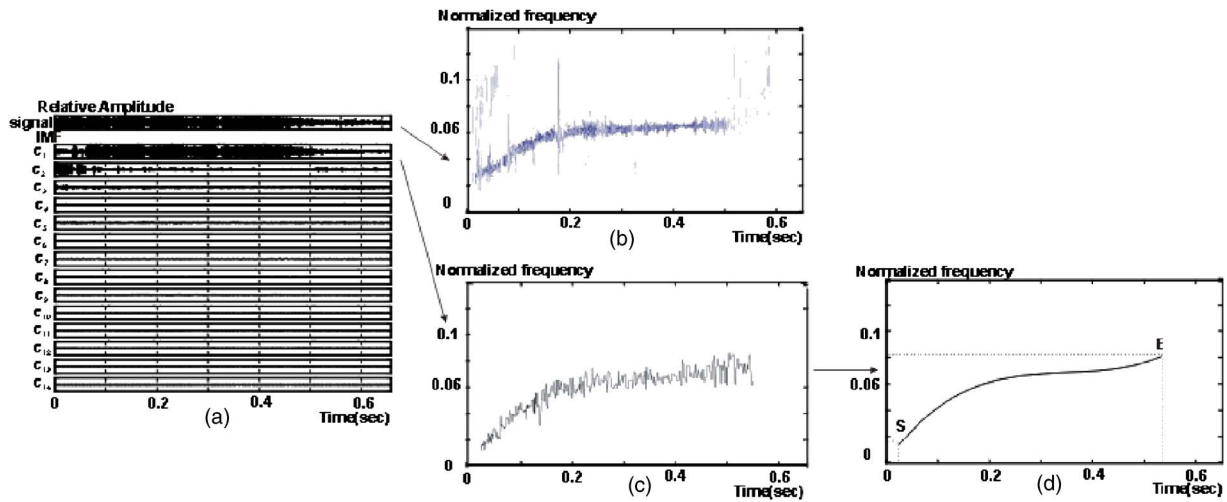
FIG. 4. (Color online) Illustration of the method for the killer whale vocalization analysis. (a) Empirical mode decomposition of the original signal. (b) Time-frequency representation of the original signal. (c) Denoising: time-frequency representation of the first mode. (d) Features extraction: three-order polynomial interpolation (S and E: start and end of the vocalization).

clicks, with various amplitudes and variant rhythm. This intrinsic variability in time domain is shown in Figs. 5(a), 5(b1), and 5(b2). Figure 5(a) shows one hundred clicks in a sperm whale creak.

The Fourier spectrogram, the wavelet transform and the HHT were applied. These time-frequency representations were used for detecting the click presence in the creak—by using a simple threshold or by pinpointing the spectrum energy maximum.

Table I gives results obtained from the spectral energy. From the Fourier spectrogram, this energy maximum was calculated

$$\max_l \sum_{k=k_0}^{K} S_w^2 \left[\frac{l}{F_s}, k \right] \tag{8}$$

with $S_w = |fft(\cdot)|$, l is the number of the sample ($l = 0, \dots, N - 1$ with N the length of the click), k_0 corresponds to $0.05 F_s$ and K to $0.3 F_s$. [Energy maximums from the HHT and from the wavelet transform were, respectively, calculated with Eqs. (6) and (7)].

We calculate the maximum of energy when the click is present, called $En(click)$ and when the click is absent (only noise), called $En(noise)$. We focus on the ratio between $En(click)$ and $En(noise)$. Results are grouped according to signal-to-noise ratio (SNR). Note that high SNR are obtained with clicks present in sequences of regular clicks and low SNR with clicks present in creaks.

In Table I, in regular clicks, $En(noise)$ is inferior to 1, showing that noise is weak in this specific recording. The three methods result in large ratios between $En(click)$ and $En(noise)$. With a ratio over 20:1, a detector based on a threshold could be envisioned. Also, when $\text{SNR} \approx 20$ dB, the Fourier spectrogram could be sufficient for detecting clicks.

For analyzing creaks, it could be more complex. As mentioned above, sperm whales modulate the amplitude of their clicks during creaks. The clicks detector could be efficient if the ratio between $En(click)$ and $En(noise)$ is still over 20:1. From $\text{SNR} \approx 10$ dB, we obtained a ratio of 6.5:1 for the

Fourier spectrogram (it was three times less than with regular clicks) while when applying the wavelet transform and the HHT, this ratio is still high ($>25:1$).

When the SNR decreases further, distinguishing the signal parts when clicks are present from signal parts when only noise is present becomes more difficult when using a threshold based on the Fourier spectrogram alone. The ratio is lower than 3:1 when SNR is equal to 5 dB. This is not the case with the wavelet transform and the HHT which are more resistant when SNR decreases. We note that underwater noise is nonuniformly distributed on all IMF and by choosing only the first IMF, we can easily reduce its impact (0.07) and keep a sufficient ratio (36.3:1) between $En(click)$ and $En(noise)$.

The higher the ratio, the more justified the use of a threshold for detecting clicks.

IV. DISCUSSION

While the Fourier spectrogram is the classical estimator for providing a time-frequency representation and analyzing signals emitted by whales, the HHT has some significant advantages: The decomposition does not need any vector basis, the decomposition is finite, and it is possible to provide a time-frequency representation.

Like with other methods of decomposition (Fourier, wavelet), EMD allows the filtering of the original signal and the focusing on the range of whale frequencies. Note that in certain cases, when noise perturbs the calculation of the instantaneous frequency ($\text{SNR} < 0$ dB and noise frequency band overlap with signal frequency band), another time-frequency representation applied on the IMF could be favored to the Hilbert transform (Janosi and Muller, 2005).

For signal analysis, the Fourier spectrogram requires the use of a time window. The features of this time window influence time-frequency representation; a compromise has to be set between focusing on the time resolution and on the frequency resolution. When frequency changes during recordings, this compromise must vary as well; consequently
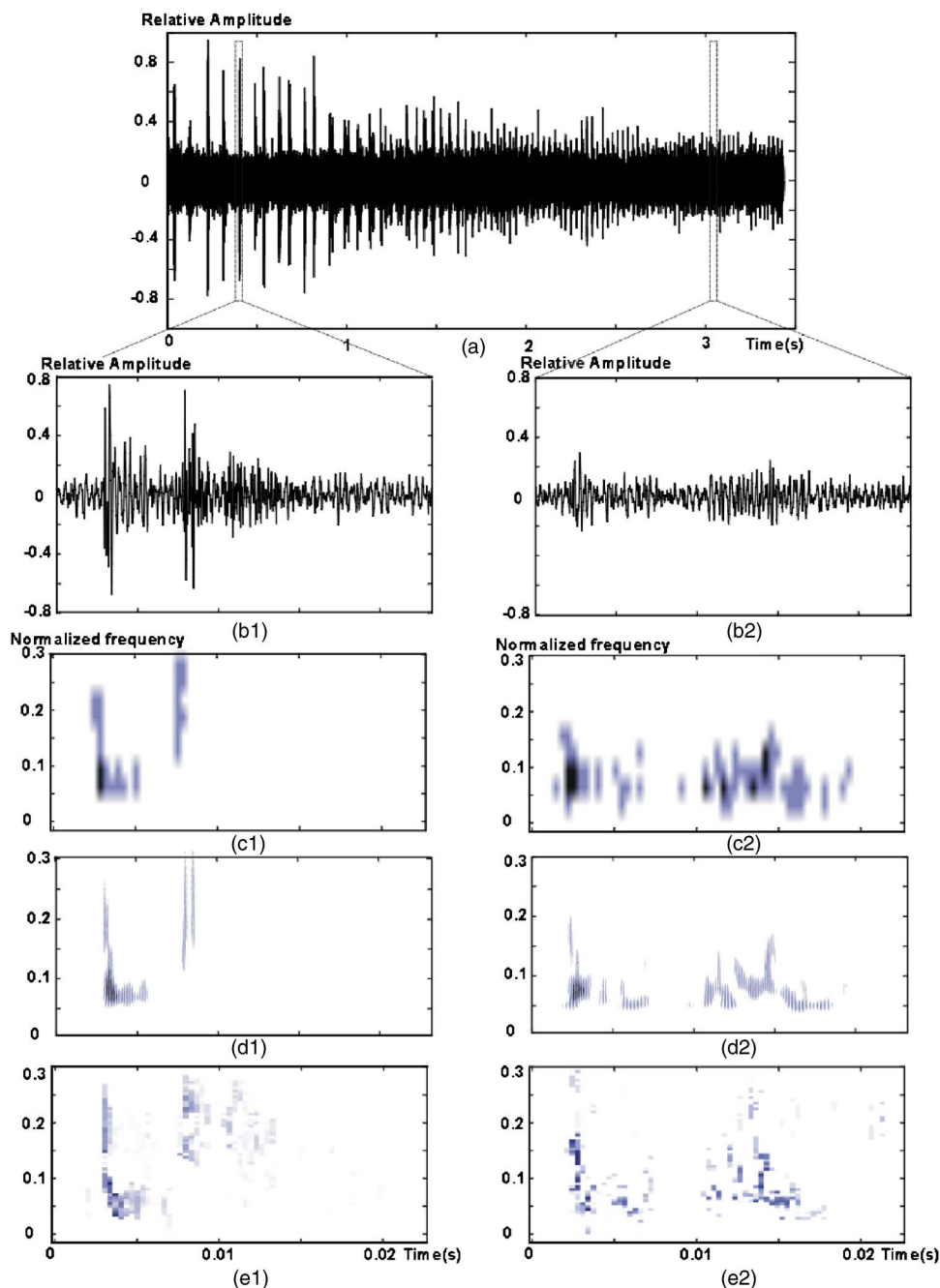
FIG. 5. (Color online) Comparison of sperm whale creak detection. (a) Sperm whale creak. (b1–2) Time-signal of sperm whale clicks. (c1–2) Fourier spectrogram. (d1–2) Morlet wavelet transform. (e1–2) The HHT.

the fixed length for this time window is not the optimal choice. Contrary to the Fourier spectrogram, the time window used for the HHT does not impact the frequency resolution.

The HHT has one more advantage: the time-frequency representation is the plot of the instantaneous frequency. This

means that we do not have to choose a specific threshold for obtaining this binary time-frequency representation.

Moreover, the time-frequency representation can be analyzed by considering instantaneous frequencies calculated from each IMF, such as the one-dimensional (1D) signals. Step 3 becomes easier: To extract any kind of information

TABLE I. Energy maximum of the three time-frequency representations. *En*(*click*) and *En*(*noise*) are the measurements of the energy maximum, respectively, in the presence of click or only of noise. Mean calculated on 20 clicks extracted from regular clicks, 20 clicks from creaks with SNR ≈ 10 dB, and 20 clicks from creaks with SNR ≈ 5 dB.

Clicks from	SNR	Fourier spectrogram			Wavelet transform			HHT		
		En(*click*)	*En*(*noise*)	Ratio	*En*(*click*)	*En*(*noise*)	Ratio	*En*(*click*)	*En*(*noise*)	Ratio
Regular Clicks	20 dB	15.55	0.78	19.9:1	56.30	0.81	69.5:1	17.52	0.18	97.3:1
Creaks	10 dB	11.99	1.85	6.5:1	47.74	1.78	26.8:1	8.22	0.22	37.4:1
Creaks	5 dB	4.15	1.56	2.7:1	25.02	1.43	17.5:1	2.54	0.07	36.3:1

from the time-frequency representation, it is not necessary to define a complex 2D post-treatment as is often done. In (Hory *et al.*, 2002; Mellinger, 2005), post-treatments require contour models for contour extractions of vocalization patterns. These kinds of 2D algorithms could be an obstacle for real time application. With the HHT time-frequency representation, we directly access the change of frequencies, as illustrated in Fig. 4. We can focus this treatment on a specific IMF which contains the pertinent information—for example, only the first five modes are considered when regular sperm whale clicks have to be analyzed (Adam, 2006). With each monocomponent vocalization, only a single IMF can be used for extracting different characteristics. The evolution of frequency obtained from the Hilbert transform of this IMF can be parametrized by a polynomial fit—choosing a three-order polynomial regression is sufficient for the example shown in Fig. 4(d)—facilitating the compression and the extraction of the classical information, such as the vocalization duration and the start and end of the vocalization frequency, etc. Thus this information can be used for classification, as input of a decision system, such as an artificial neural network, for example.

This is perhaps the most important advantage of the HHT.

Regarding the transient signals emitted by whales, such as sperm whale clicks, the wavelet transform and the HHT show better robustness than the Fourier spectrogram when the SNR decreases. This property could be used with success for detectors based on a threshold.

Because the clicks are transient and of the impulse type, the wavelet transform and the HHT are attractive alternatives to the Fourier estimator. The coefficients of the wavelet transform could be used for defining an acoustic signature (Lopatka, 2005). By isolating the maximum of the instantaneous frequency of only the four first modes, the HHT analysis contributes to this objective.

V. CONCLUSION

This paper presents an application of the Hilbert Huang transform applied to bioacoustics signals, and particularly vocalizations and clicks emitted by whales, present in continuous recordings. This technique contributes to characterizing these kinds of signals, and could be used successfully in the definition of detectors and/or classifiers.

In this paper, we have compared three different time-frequency representations (Fourier spectrogram, Morlet wavelet transform, and the HHT). With the calculation of the instantaneous frequency, the Hilbert transform allows for the concentration of frequency energy and the avoidance of the uncertainty principle.

As with the other methods, it is possible to define filters based on the HHT, i.e., the analysis could be done with certain specific modes resulting from the decomposition. Retaining one or few IMF is done to denoise the original signal. This approach also corresponds to retaining only the modes containing the pertinent information. This technique could contribute to distinguishing two whales which emit signals simultaneously.

Finally, it is possible to analyze the frequency variation of each IMF with 1D technique and not 2D approaches, as is commonly done for extracting information from the time-frequency representations.

Adam, O. (**2006**). "The use of the Hilbert Huang transform to analyze Transient Signals Emitted by Sperm Whales," *Applied Acoustics*, **67**(11–12), 1061–1070 (Elsevier, New York).

Altes, R., A. (**1980**). "Detection, estimation, and classification with spectrograms," J. Acoust. Soc. Am. **67**, 1232–1246.

Barlow, J., and Shannon, R. (**2004**). "Estimates of the percentage of sperm whales missed on combined visual and acoustic surveys in the eastern Pacific Ocean," J. Acoust. Soc. Am. **116**(4), 2588.

Boashash, B. (**2003**). *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference* (Elsevier, New York).

Cohen, L. (**1995**). *Time-Frequency Analysis* (Prentice-Hall, Englewood-Cliffs, New Jersey).

Dalheim, M. E., Fisher, H. D., and Schempp, J. D. (**1984**). "Sound production by the grey whale and ambient noise levels in Laguna San Ignacio, Baja California Sur, Mexico," in *The Gray Whale Eschrichtius Robustus*, edited by M. L. Jones, S. L. Swartz, and S. Leatherwood (Academic Press, Orlando), pp. 511–541.

Erbe, C. (**2000**). "Detection of whale calls in noise: Performance comparaison between beluga whale, human listeners and neural network," J. Acoust. Soc. Am. **108**(1), 297–303.

Fargues, M. P., and Bennett, R. (**1995**). "Comparing wavelet transforms and AR modeling as feature extraction tools for underwater signal classification," asilomar, *29th Asilomar Conference on Signals, Systems and Computers*, Vol. **2**. p. 915.

Flandrin, P., Rilling, G., and Goncalves, P. (**2003**). "Empirical Mode Decomposition as a filter bank," IEEE Signal Process. Lett. **11**(2), 112-114.

Gillespie, D. (2004). "Detection and classification of right whale class using an edge detector operating on smoothed spectrogram," J. Can. Acoust. **32**(2), 39–47.

Goupillaud, P., Grossmann, A., and Morlet, J. (**1984**). "Cycle-octave and related transforms in seismic signal analysis," Geoexploration, **23**, 85–102.

Hory, C., Martin, N., and Chehikian, A. (**2002**). "Spectrogram segmentation by means of statistical features for non-stationary signal interpretation," IEEE Trans. Signal Process. **50**(12), 2915–2925.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (**1998**). "The empirical mode decomposition and the Hilbert transform spectrum for nonlinear and nonstationary time series analysis," Proc. R. Soc. London **454**, 903–995.

Huele, R., and Udo De Haes, H. (**1998**). "Identification of individual sperm whales by wavelet transform of the trailing edge of the flukes," Marine Mammal Sci. **14**, 143–145.

Ioana, C., and Quinquis, A. (**2004**). "On the use of time-frequency warping operators for analysis of marine-mammals signals," *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), 2, 605–608.

Janosi, I. M., and Muller, R. (**2005**). "Empirical mode decomposition and correlation properties of long daily ozone records," Phys. Rev. B **71**, 056126–1-5.

Jarvis, S., Moretti, D., Morrissey, R., and DiMarzio, N. (**2003**). "Passive monitoring and localization of marine mammals in open ocean environments using widely spaced bottom mounted hydrophones," J. Acoust. Soc. Am. **114**(4), 2405–2406.

Kadambe, S., and Orr, R. (**1995**). "Application of Gabor analysis for detection, estimation and classification of underwater acoustic data," asilomar, *29th Asilomar Conference on Signals, Systems and Computers*, Vol. 2, p. 569.

Kuperman, W. A., D'Spain, G. L., and Heaney, K. D. (**2001**). "Long range source localization from signal hydrophone spectrograms," J. Acoust. Soc. Am. **109**(5), 1935–1943.

Lewicki, M. (**2002**). "Efficient coding of natural sounds," Nat. Neurosci. **5**(4), 356–363.

Lopatka, M., Adam, O., Laplanche, C., Zarzycki, J., and Motsch, J. F.

(**2005**). "An attractive alternative for sperm whale click detection using the wavelet transform in comparison to the Fourier spectrogram," Aquat. Mamm. **31**(4), 462–466.

Matthews, J., Rendell, L. E., Gordon, J., and Macdonald, D. W. (**1999**). "A review of frequency and time parameters of cetacean tonal calls," Bioacoustics 10, 47–71.

Mellinger, D. (**2005**). "Automatic detection of regularly repeating vocalizations," J. Acoust. Soc. Am. **118**(3), 1940.

Meyer, Y. (**1993**). *Wavelets: Algorithms and Applications* (SIAM Press, Philadelphia, PA).

Nakamura, K., and Akamatsu, T. (**2004**). "Comparison of click characteristics among odontocete species," in *Echolocation in Bats and Dolphins*, edited by Thomas, Moss, and Vater (University of Chigaco, Chicago), pp.

36–43.

Potter, J. R., Mellinger, D. K., and Clark, C. W. (**1994**). "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262.

Oswald, J. N., Barlow, J., and Norris, T. F. (**2003**). "Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean," Marine Mammal Sci. **19**(1), 20–37.

Tiemann, C. O., and Porter, M. (**2003**). "Automated model-based localization of sperm whale clicks," *Proceedings of Oceans Conference*, **2**, 821–827.

Wahlberg, M. (**2002**). "The acoustic behaviour of diving sperm whales observed with a hydrophone array," J. Exp. Mar. Biol. Ecol. **281**, 53–62.

Semi-automatic classification of bird vocalizations using spectral peak tracks

Zhixin Chen[a)] and Robert C. Maher[b)]

*Department of Electrical and Computer Engineering, Montana State University,
Bozeman, Montana 59717-3780*

Automatic off-line classification and recognition of bird vocalizations has been a subject of interest to ornithologists and pattern detection researchers for many years. Several new applications, including bird vocalization classification for aircraft bird strike avoidance, will require real time classification in the presence of noise and other disturbances. The vocalizations of many common bird species can be represented using a sum-of-sinusoids model. An experiment using computer software to perform peak tracking of spectral analysis data demonstrates the usefulness of the sum-of-sinusoids model for rapid automatic recognition of isolated bird syllables. The technique derives a set of spectral features by time-variant analysis of the recorded bird vocalizations, then performs a calculation of the degree to which the derived parameters match a set of stored templates that were determined from a set of reference bird vocalizations. The results of this relatively simple technique are favorable for both clean and noisy recordings. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2345831]

I. INTRODUCTION

Several engineering applications require real time identification of birds while in flight, foraging, or roosting. These include systems to help avoid collisions between birds and aircraft (Pascarelle *et al.*, 2004), systems to track migratory birds in the vicinity of wind turbine generators (NWCC, 2004), and ornithological measurements systems to help understand avian behavior and migratory patterns, particularly at night and in unfavorable meteorological conditions. Among the possible means to identify the bird species are their vocalizations. Thus, there exists a need for research in on-line acoustical bird classification systems capable of running unattended and in real time.

In recent years many off-line techniques for classifying bird species based on recorded vocalizations have been proposed and developed. The most successful techniques are based on manual inspection and labeling of bird sound spectrographs by experts, but this process is tedious and dependent upon the subjective judgment of the observer (Kogan and Margoliash, 1998). The reliability of classification can be improved if a panel of experts is used, but this is expensive, time consuming, and unsuitable for real time classification. Nevertheless, the fact that the manual inspection of sound spectrographs tends to yield correct judgments has encouraged research into automatic classification using objective standards derived from expert opinions.

Several of the existing automatic off-line bird vocalization classification techniques are based on traditional speech recognition methods (Rabiner and Juang, 1993). Anderson, Dave, and Margoliash (1996) used dynamic time warping (DTW) for automatic recognition of birdsong syllables from continuous recordings. Their method directly compared the spectrograms of input bird sounds with those of a set of predefined templates representative of categories chosen by the investigator. They applied this method to vocalizations from two bird species recorded in a low noise environment and achieved 97% accuracy in the syllables of stereotyped songs and 84% accuracy in the syllables of plastic (variable or indistinct syllable) songs. The method did not use amplitude normalization, so the results may be sensitive to amplitude differences. Kogan and Margoliash (1998) used DTW and hidden Markov models (HMM) to classify bird sounds based on the syllables extracted from continuous recordings. Their method began by extracting linear predictive coding (LPC) coefficients or mel-frequency cepstral coefficients (MFCC) from a set of bird syllables and then used DTW or HMM for recognition. This method was found to perform well for two specific birds in a low noise environment. The method worked less well in noisy environments or with short duration bird vocalizations. Ito, Mori, and Iwasaki (1996) extracted two time-varying spectral features from syllables and used dynamic programming (DP) matching to classify budgerigar contact calls, and found that the method performed well. However, only the frequencies of the spectral features were used, not the spectral powers, so the method may not be appropriate for other bird sounds with different spectral structure.

Other researchers have developed classification methods specifically tailored to bird vocalizations. McIlraith and Card (1997) conducted research on the recognition of songs of six bird species. In their method the bird songs were represented with spectral and temporal parameters of the songs. They reduced the complexity of the search space by selecting features exhibiting the greatest discrimination, then used a neural network for classifying the bird songs. Their method achieved good performance but the neural classifier required

---

[a)]Electronic mail: chen@montana.edu
[b)]Corresponding author. Electronic mail: rob.maher@montana.edu

considerable computation. Härmä (2003) proposed an alternative method for bird sound classification. He observed that for many songbird vocalizations, a large class of syllables can be approximated as brief sinusoidal pulses with time varying amplitude and frequency. Although this model is too simple for certain bird sounds, the system provided good recognition results for species with tonal vocalizations. In a subsequent study, Härmä and Somervuo (2004) classified bird sounds into four classes based on their harmonic structure, where each harmonic component was modeled with one time-varying sinusoid. No classification statistics were reported, but they found that the signal models appropriately represented the spectral structure of 93% of the syllables in their database.

In summary, the existing methods are well suited to their specific application, but they also have some limitations. The DTW and HMM techniques did not perform well in noisy environments or for bird sounds with short duration and variable amplitude. The neural classifier required a very high computational complexity. The use of one or two spectral peak tracks is appealing for its simplicity and robustness to noise. Thus, in this paper we describe a spectral pattern detection method with relatively low computational complexity (i.e., suitable for a final implementation in real time) that can be used to classify in real time tonal bird vocalizations (harmonic or inharmonic) in the presence of realistic background noise levels.

The remaining sections of this paper are organized as follows. First, we review briefly the basic structure and terminology of bird vocalizations. We then describe the rationale and features of the proposed semi-automatic classification method, including our simulation results and interpretation, and compare the results to classification based on conventional DTW and HMM methods. Finally, we conclude with a summary and several suggestions for future research in this area.

II. BIRD SOUNDS

Birds are able to produce a wide variety of sounds. Air from the lungs is forced through the bronchi to the syrinx, which is the major source of vibratory modulation. Sound from the syrinx passes through the resonant structures of the trachea, larynx, mouth, and beak.

Bird vocalizations can be divided into the general categories of elements, syllables, phrases, calls, and songs (Krebs and Kroodsma, 1980; Catchpole and Slater, 1995). *Elements* can be regarded as the elementary sonic units in bird vocalizations. A *syllable* includes one or more elements and is usually a few to a few hundred milliseconds in duration. *Phrases* are short groupings of syllables. *Calls* are generally compact sequences of phrases, while *songs* are long and complex vocalizations. An example of the hierarchy is shown in Fig. 1. The reader is cautioned to be aware that the details of each category often show individual and geographic variations and temporal plasticity, even within a single species (Kroodsma and Miller, 1996; Krebs and Kroodsma, 1980). Thus, a classification strategy consisting of a simple tone sequence detector tuned to match a specific
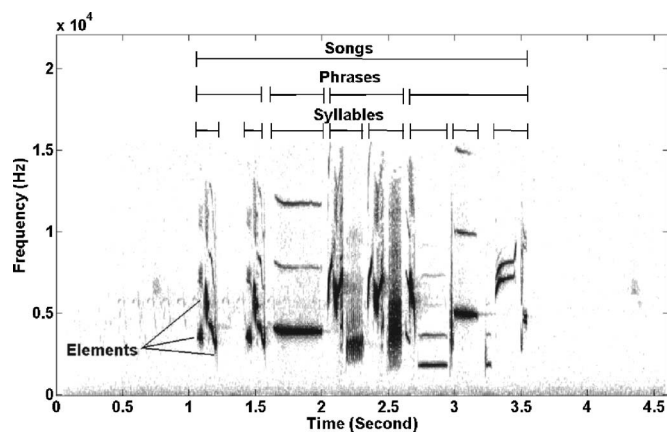
FIG. 1. Descriptive hierarchy of bird vocalization. The bird vocalization can be divided into four hierarchical levels: element, syllable, phrase, and song (or call). The classification experiment reported in this paper operates at the syllable level in the hierarchy.

example bird song is unlikely to be useful in engineering applications due to natural variation in the song details.

Syllables can range from being nearly tonal whistles, to harmonic sounds with a distinct fundamental frequency, to inharmonic bursts, or even to being noise like (Nowicki, 1997; Kahrs and Avanzini, 2001; Fagerlund, 2004). In the case of voiced harmonic sounds, the fundamental frequency range typically lies between 500 and 5000 Hz. The spectral content of a harmonic sound can also vary with time as the bird changes the shape and length of the vocal tract and beak. Figure 2 shows examples from songs and calls from different species, illustrating only a small variety of the complex sounds birds can produce.

III. CLASSIFICATION METHOD AND EXPERIMENTAL RATIONALE

The bird sound classification problem is similar to many existing pattern detection and classification problems. The classification procedure described in this paper follows the standard design cycle of Duda *et al.* (2001), as depicted in Fig. 3.

The *data collection* process included the acquisition of audio recordings using recording apparatus similar to what is anticipated in a real time classification system. The data were obtained in the field with a variety of ambient noise sources such as wind, motor traffic, and aircraft.

The *feature choice* process was based on our examination of the spectrotemporal characteristics of the bird sounds of interest. This examination was supplemented by a review of prior research into bird sound recognition and our prior experience with a variety of signal processing methods useful for data extraction in speech, environmental sounds, and music. For the purposes of classification in the presence of background noise, we have found that subsyllable elements are difficult to use in practice simply due to the difficulty in extracting reliable signal parameters for such short segments. Meanwhile, ornithologists report that higher level phrases, calls, and songs often contain regional and individual variations that will add additional degrees of freedom to the classification problem (Krebs and Kroodsma, 1980). This led us
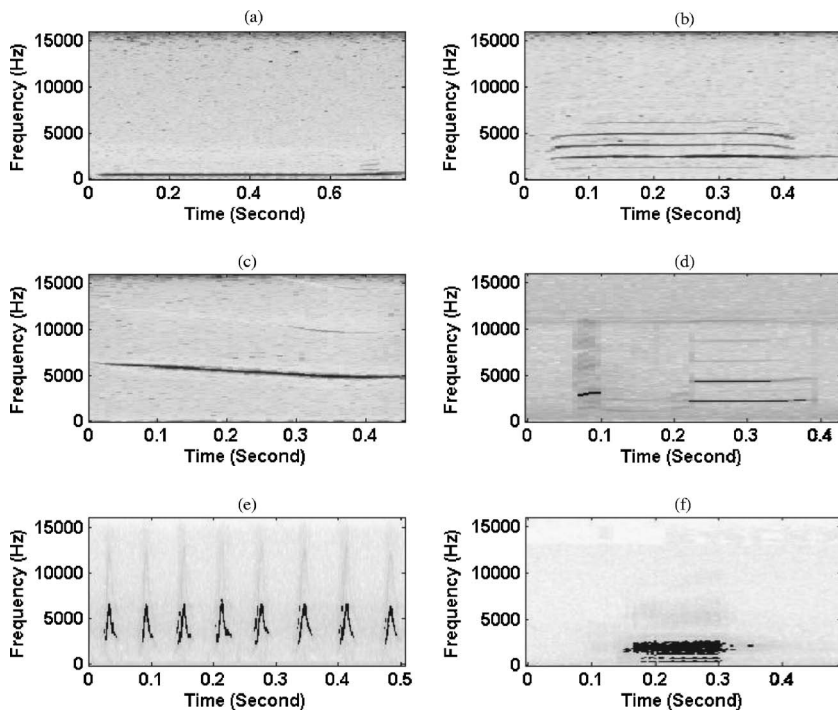
FIG. 2. Example time-variant spectra of bird sounds from several bird species. (a) Mourning dove (*Zenaida macroura*), (b) Red-tailed hawk (*Buteo jamaicensis*), (c) Red-winged blackbird (*Agelaius phoeniceus*), (d) Herring gull (*Larus argentatus*), (e) Brown-headed cowbird (*Molothrus ater*), (f) Mallard (*Anas platyrhynchos*).

to focus on syllable-level classification as the starting point for our proposed method. Restricting the process to single syllables allows the problem to be tractable, but a fully practical automatic classification system will need to accommodate the many different syllables associated with a particular bird species. Nevertheless, the characteristic strengths and weaknesses of our proposed classification method can be assessed using a single syllable study.

We observed the patterns and trends in our bird sound database and developed a set of discrimination parameters including spectral frequencies, frequency differences, track shape, spectral power, and track duration. The resulting framework uses measurements of the principal peak tracks in each sound syllable based on our observations and measurements of the recorded sound set.

The classifier *model choice* and *training* processes (de-

scribed in the following section) consisted of selecting a representative syllable from each desired bird species in the database, deriving the features, or discrimination parameters, according to the basic peak track model, and then determining an allowable discrepancy between the derived features and the data for which a match would still be allowed.

The *evaluation* process, described in Sec. V, consisted of an iterative procedure comparing the error rate of the proposed classifier under a range of parameter range adjustments and additive noise conditions.

IV. DESCRIPTION OF THE SPECTRAL PEAK TRACKS METHOD

As mentioned above, spectral features have been found to be useful for bird syllable classification. Specifically, the use of spectral peak tracks provides a compact, distinctive, and computationally tractable basis for classification (Härmä, 2003). A spectral peak track is formed by segmenting the input signal into a set of overlapping short-time frames, calculating the Fourier transform magnitude for each short-time frame, and then matching peaks from one spectral magnitude frame to the next. If the input signal contains underlying sinusoidal partials, the resulting spectral peaks will persist for several successive frames. The corresponding peaks (e.g., peaks nearest in frequency on successive frames) are identified and linked from one frame to the next to form a connected sequence, known as a *peak track*, indicating the amplitude and frequency trajectory of the underlying partial. The vocal sounds of many birds are found to be well modeled by the peak track model (Härmä and Somervuo, 2004).

We extend the basic peak track method by including a variable number of peak tracks, which allows tonal, harmonic, or inharmonic combinations. We also enhance the
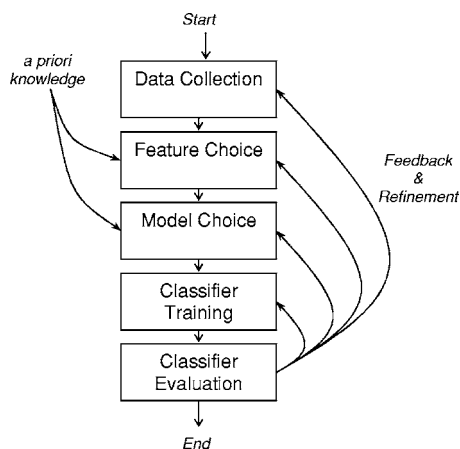
FIG. 3. Pattern classification design cycle (Duda *et al.*, 2001). Although numerous pattern matching procedures have been proposed and demonstrated, no universal system yet exists. Nonetheless, the design cycle depicted here is a useful common starting point for classification system design.

robustness of the peak track method by encoding the power and shape of the spectral peak tracks to capture the additional characteristic behavior of each vocalization.

In the proposed method we extract a variable number of spectral peak tracks from one syllable of the desired bird sound. The number of peak tracks selected is determined by the relative power of the tracks from the entire syllable. A set of descriptive parameters is then derived from the selected spectral peak tracks. These parameters include the frequencies, frequency differences, track shape, relative powers, and the track duration. We have found these parameters to work well in the presence of moderate noise and competing background sounds. The method also has low computational complexity and ease of implementation.

In the following sections we describe the three major steps of the spectral peak tracks method: spectral peak tracks search, feature extraction, and target and recognition.

A. Spectral peak tracks search

Step A1: The first processing step is to determine the spectral peak tracks of the syllable. A digital recording of the syllable with 16-bit samples and a 16 kHz sample rate is first high pass filtered with a 100 Hz cutoff in order to remove the low frequency background noise attributable to wind or mechanical sounds. The filtered signal is then segmented into Hamming windowed 256-sample frames with 128-sample overlap (8 ms frame hop). This is a reasonable frame rate based on the expected spacing of the spectral peaks. The *raw spectral representation* (i.e., the short-time discrete Fourier transform) is obtained with a Fast Fourier Transform (FFT) algorithm.

15 frames (120 ms) of the sample recording from the time interval prior to the start of the vocalization are used to estimate the background noise level. The average noise level for the 15 frames is calculated and used to set the amplitude threshold for the subsequent peak track detection. The estimated background noise level is also used to clarify the onset and release boundary of the syllable, and to distinguish the temporal boundaries of syllables with two or more parts. A simple energy-based detection algorithm is used to identify syllables with more than one part. If more than one part is found, only the part with the largest energy is retained, but the multipart detection is noted for use during the matching and recognition process.

Step A2: Next, the spectral peak tracks are derived using the McAulay and Quatieri (MQ) procedure (McAulay and Quatieri, 1986; Smith and Serra, 1987; Ellis, 2003). We refer to this step as the *coarse search*. In each frame the magnitude of the FFT data is examined to locate peaks in the spectrum. A peak is identified by three adjacent spectral magnitude coefficients within a frame where the middle coefficient is larger than both its higher and lower frequency neighboring coefficients, and the magnitude also exceeds the noise threshold determined in Step A1. This is done to remove peaks with very weak magnitude that are likely to be caused by background noise or by sidelobes of the Hamming window spectrum. We use a quadratic fit for the three spectral coefficients defining the peak in order to refine the frequency and magnitude estimate for the peak (Smith and Serra, 1987).

The refined peaks in one frame are compared to the peaks in the subsequent frame, and those spectral peaks that match well from one frame to the next are connected to form candidate peak tracks. A good peak-to-peak match is determined by locating a peak in the subsequent frame that is closest in frequency to each peak in the current frame. Acceptable matches are also restricted by a maximum rate of change in frequency corresponding to the most rapid frequency sweep found in actual bird syllables. We have found that a 200 Hz range works well for spectral peaks with frequencies above 2 kHz, while a 100 Hz range is used as the allowable frequency difference for peaks below 2 kHz. In some cases there will not be a suitable matching peak in the subsequent frame, while in other cases there may be several possible matches. Any conflicts are resolved by finding the match that minimizes the frequency and magnitude difference between the tentatively matched peaks. If no good match is found, we assume that the current peak is not continuous with any of the existing peak tracks, and the peak track associated with that peak is marked as "dead." Similarly, if a peak in the subsequent frame is not found to be a match for any of the peaks in the current frame, the new peak is considered as the "birth" of a new peak track. Thus, the coarse search process continues for each frame of spectral data, creating sets of candidate peak tracks.

Step A3: The spectral peak tracks produced by the coarse search in Step A2 may be discontinuous, too short, or otherwise inconsistent and poorly constructed. Therefore, a *fine search* is conducted to seek a higher level of structure in the syllable peak track data. We first identify any brief gaps (three or fewer frames) in the candidate peak tracks and reconnect across the gap under the assumption that the gap is due to noise or thresholding. Next, we prune away tracks that are inconsistent with the expected fundamental frequency range of the bird vocalization, such as tracks below 500 Hz. We then eliminate any closely spaced tracks, keeping the track with the larger magnitude, again under the assumption that the weaker track is due to spurious noise. Finally, we calculate the mean or median frequency, track length, and track power for every peak track and store the results of the fine search in the track analysis database.

Step A4: If the track list derived for the syllable has only one spectral peak track, the peak track search is terminated. If the track list includes two or more tracks, a second fine search is conducted. First, any relatively weak peak tracks are eliminated from the track list, thereby retaining only the tracks with the greatest signal to noise ratio. These strong peak tracks are sorted according to power and duration. If any of the strong peak tracks are found to be much shorter or longer in duration than the strongest peak track, or if the onset or offset of one of the weaker tracks is considerably different from the strongest peak track, those marginal tracks are eliminated from the track list. The resulting list of primary peak tracks (usually eight or fewer tracks) is retained for use in the matching procedure.

An example of the primary peak tracks for a harmonic syllable is shown in Fig. 4. The syllable has five spectral

FIG. 5. Frequency range and shape example. (a) Peak track representation of a syllable from bird species I, (b) Peak track representation of a syllable from bird species II. The classification algorithm compares both the range (frequency extent) and the shape of the spectral peak tracks to improve the reliability of the syllable classification.

FIG. 4. Result of the spectral peak tracks search. (a) Frequency versus time representation (spectrogram) of one syllable of a red-tailed hawk (*Buteo jamaicensis*), (b) Spectrogram showing the search results from the spectral peak track method, (c) A three-dimensional representation of the spectrogram (amplitude versus frequency versus time).

peak tracks. Among these tracks, Track D is the most significant, Track B is the next most significant, followed by Track C, Track A, and finally Track E. Since Tracks A and E are sufficiently weak to be immersed in the background noise, they are not retained in the search phase. Consequently, only Tracks D, B, and C are selected in the spectral peak tracks search phase.

B. Feature extraction

Once the primary spectral peak tracks are determined, we analyze the track contours to obtain a set of descriptive parameters. The parameters include the frequencies, the frequency differences, the relative power, the shape, and the duration of the spectral peak tracks. These parameters were selected based on several key insights. First, the peak track frequencies are observed to be distinctive for many of the bird species in our recorded database, and are therefore considered useful for pattern classification (see the example in Fig. 5). Second, we recognize that there are different bird species with syllables in overlapping spectral ranges, meaning that calculating only the mean frequency will be insuffi-

cient to distinguish between the species. In some cases the presence of several peak tracks with specific frequency relationships can be a key discriminator between species, while in other cases the shape and trend of the peak tracks may be the most important distinguishing feature. Finally, it may be that two different syllables are otherwise similar in frequency range, track frequency spacing, and track shape. In these cases we have found that comparisons between the relative power and the total duration of the tracks are necessary to obtain a reliable discrimination between the species.

For each primary track we first calculate the frequency at the start, end, midpoint, first quartile and third quartile point, which are F_1, F_2, F_3, F_4, and F_5 respectively. Next, we calculate the difference between the mean frequencies of the three primary tracks, which we denote as FD_1 and FD_2. Note that if there is only one primary track in the syllable, both FD_1 and FD_2 will be set to the mean frequency of the track, while if there are only two tracks in the syllable, both FD_1 and FD_2 will be set to the difference between the mean frequencies of these two tracks.

Next, a simple calculation is performed to describe the *shape* of each track. The shape is described by three parameters, S_1, S_2, and S_3. The first parameter, S_1, is expressed as

$$S_1 = \frac{F_3 - F_1}{\text{Frame\_Num\_1}}, \tag{1}$$

where F_1 is the starting frequency, F_3 is the midpoint frequency, and Frame_Num_1 is the number of frames be-

TABLE I. Output parameters for bird syllables. The first column represents the number of syllable parts and peak tracks. The second column represents the twelve parameters for each peak track.

Overall description	For each track
Number of parts in the syllable	Starting frequency (F_1)
Number of peak tracks	Ending frequency (F_2)
	Middle frequency (F_3)
	1st quartile frequency (F_4)
	3rd quartile frequency (F_5)
	Frequency difference (FD_1)
	Frequency difference (FD_2)
	Shape (S_1)
	Shape (S_2)
	Standard deviation (S_3)
	Relative power
	Duration

tween the start point and the middle point of the peak track. Similarly, the second parameter, S_2, is expressed as

$$S_2 = \frac{F_3 - F_2}{\text{Frame\_Num\_2}}, \quad (2)$$

where F_2 is the ending frequency and Frame_Num_2 is the number of frames between the endpoint and the middle point of the peak track. The third parameter, S_3, is the standard deviation of the frequency of the track calculated over the entire duration.

Finally, we also calculate the relative power of each track by determining the fraction of the total power in each frame due to the current track's peak, then averaging these power fractions over the length of the track.

The 12 parameters derived for each peak track have different units and magnitudes. Therefore, the parameters are normalized so that the parameter vector can be used in a distance calculation for classification. Of course, any number of mathematical functions could be invented for the distance calculation. We have developed a set of empirical coefficients that balance the variation of each dimension: scaling the frequency and frequency difference parameters by 62.5, scaling the shape parameters by 1.6, scaling the power by 2, and scaling the duration by 125 frames provides this balance.

The final step for feature extraction is to collect the derived parameters into a data structure ready for the subsequent matching process. Table I shows the output parameter summary.

C. Target and recognition

The target phase is conducted with representative vocalizations for each bird species to be tested. The representative vocalizations are intended to represent the desired syllable in the best possible manner, so the selection is conducted manually. The spectral peak tracks and features are extracted for each representative vocalization, creating a template for each bird syllable in the test set. These syllable templates then become the target database for the matching process, as described next.

The recognition phase begins with a recorded segment of an unknown bird sound. In our study the sound is assumed to be edited manually to contain a suitable segment for the matching process.

Step C1: Determine the number of syllable parts and peak tracks and store them in the output parameter list for the unknown sound.

Step C2: Calculate the *frequency distance* and the *number of syllable parts* match between the unknown sound and each template in the target database. The first five parameters (frequencies F_1 through F_5) are used in this step because in our experience the frequency information has been found to be the most important parameter of the matching process. The strongest peak track from the template is compared to each peak track in the test sound since the strongest peak track may differ from one example to the next even within the same species. If the number of syllable parts in the test sound differs from the number of parts in a template, the frequency distance is increased as a "penalty" to account for the lower likelihood of a match.

Step C3: (a) If the minimum distance between the unknown sound and at least one of the target templates is less than an empirically determined threshold (e.g., 8 Hz, or approximately 1% of the typical fundamental frequency component), we assert that a match has been found and the template with the minimum distance from the unknown sound is deemed to be the recognition result. Note that this small difference is unlikely, considering the noise and level differences inherent in real signals. Otherwise, the potential target matches are sorted and the two templates with the smallest distances from the unknown sound are identified for further comparison.

(b) Conversely, if the frequency distance between the unknown sound and one of the target templates is large, e.g., the mean frequency difference is greater than 800 Hz, the target template is classified as a very poor match to the unknown sound, and the recognition process proceeds with the next potential match.

(c) If the number of peak tracks in the unknown sound differs substantially (e.g., a difference greater than four) from the target template, the target is considered a poor match and the recognition process moves onto the next target template. For example, if the unknown sound is nearly a pure tone with a single peak track while a target template contains seven peak tracks, we assume the unknown sound cannot be the same bird species as the template.

Step C4: If the process has identified several candidate target templates based on the frequency distance calculations, now all 12 parameters are used to calculate the total weighted distance between the unknown sound and the candidate templates. We have found that better matching results are obtained if a penalty is included for mismatched tracks: 10% multiplied by the absolute track count difference between the unknown sound and the target template is added to the distance calculation.

The target template with the minimum distance is deemed the best match to the unknown sound.

V. SIMULATION AND DISCUSSION

The original raw simulation database of bird vocalizations contained sounds with varying formats and signal integrity. Some of the sounds came from recordings in the vicinity of an airport and therefore contain substantial background noise, while other sounds came from archival records with extremely low noise. The original sample rates ranged from 8 to 24 kHz, and both mono and stereo recordings were part of the raw database.

To facilitate the simulation all recordings were converted to monophonic, 16 kHz sample rate, and 16 bits per sample. For the purposes of verifying the performance of the peak track algorithm, recordings containing multiple syllables were manually edited to isolate a single syllable with a gap at the head and tail of the syllable. The rationale for this editing step was that if the algorithm failed under these optimal conditions it would not make sense to treat the continuous syllable case, nor the real time parameter extraction process.

A. Test database

Twelve bird species were part of the test database. Each bird species had 20 sound files, including similar syllables from separate raw recordings. The species included Mallard (*Anas platyrhynchos*), American Crow (*Corvus brachyrhynchos*), Canada Goose (*Branta canadensis*), Baltimore Oriole (*Icterus galbula*), Common Nighthawk (*Chordeiles minor*), Killdeer (*Charadrius vociferous*), Osprey (*Pandion haliaetus*), Northern Cardinal (*Cardinalis cardinalis*), Blue Jay (*Cyanocitta cristata*), Great Horned Owl (*Bubo virginianus*), Trumpeter Swan (*Cygnus buccinator*), and Herring Gull (*Larus argentatus*). These bird species are common in North America and relevant to the aviation and wind turbine bird strike avoidance problem. With the exception of the Mallard syllables, all the sounds contained strong tonal components that are well matched to the spectral peak track model. Despite the lack of predominantly tonal components in the Mallard examples, the method still achieved satisfactory results on the database, as described below.

As an additional assessment measure, we augmented the database with 16 synthesized syllables: 5 single tones (or chirps), 5 harmonic sounds, 5 inharmonic sounds, and 1 two-part syllable. The augmented database elements, or *distracters*, are deterministically generated and therefore provide a means to determine the sensitivity of the matching procedure. Thus, the total database contained 28 different syllable sets, designated Category 0 through Category 27, with 20 syllable recordings in each set. The bird species category designations are shown in Table II.

In this study, we first evaluated the database with two conventional classification methods used in speech recognition (Rabiner and Juang, 1993), then with the peak track method described in this paper.

B. Classification by DTW and HMM

The first conventional classification system was based on linear prediction cepstral coefficients (LPCC) and dynamic time warping (DTW) methods (Rabiner *et al.*, 1978).

TABLE II. Bird species category designations. Category 0 through 11 are natural bird species. Category 12 through 27 are synthetic test signals with deterministic parameters.

Category	Description
0	Mallard (*Anas platyrhynchos*)
1	American Crow (*Corvus brachyrhynchos*)
2	Canada Goose (*Branta canadensis*)
3	Baltimore Oriole (*Icterus galbula*)
4	Common Nighthawk (*Chordeiles minor*)
5	Killdeer (*Charadrius vociferous*)
6	Osprey (*Pandion haliaetus*)
7	Northern Cardinal (*Cardinalis cardinalis*)
8	Blue Jay (*Cyanocitta cristata*)
9	Great Horned Owl (*Bubo virginianus*)
10	Trumpeter Swan (*Cygnus buccinator*)
11	Herring Gull (*Larus argentatus*)
12	Single chirp. Frequency linearly increases.
13	Single chirp. Frequency linearly decreases.
14	Single chirp. Frequency linearly increases and then decreases.
15	Single chirp. Frequency linearly decreases and then increases
16	Single tone.
17	Harmonic chirp. Frequency linearly increases.
18	Harmonic chirp. Frequency linearly decreases.
19	Harmonic chirp. Frequency linearly increases and then decreases.
20	Harmonic chirp. Frequency linearly decreases and then increases
21	Harmonic tone.
22	Inharmonic chirp. Frequency linearly increases.
23	Inharmonic chirp. Frequency linearly decreases.
24	Inharmonic chirp. Frequency linearly increases and then decreases.
25	Inharmonic chirp. Frequency linearly decreases and then increases
26	Inharmonic tone.
27	One syllable has two parts: inharmonic tone plus inharmonic chirp.

The second conventional system was based on mel-frequency cepstral coefficients (MFCC) and hidden Markov models (HMM) (Rabiner, 1989).

The conventional classifiers have been successful for speech recognition because both the linear prediction model and the cepstral model are well suited to the human speech production mechanism. Specifically, the speech spectrum is characterized by a harmonic sequence attributable to the glottis excitation (vocal fold vibrations) and a set of relatively broad resonances (formants) due to the vocal tract. As mentioned in Sec. II, bird vocalizations are similarly produced by an excitation source that is spectrally shaped by resonances of the trachea, larynx, mouth, and beak. However, the differences in the excitation signal and the generally smaller dimensions of the bird vocal structure compared to the human speech production system result in more widely spaced spectral partials and more narrowly spaced resonances in typical bird sounds compared to human speech patterns. Thus, a conventional speech recognition technique will not necessarily work well with bird sounds, and this concern is borne out in the following test results.

1. Description of the classification system based on LPCC and DTW

Bird vocalization is a time-dependent process. Two similar bird syllables may have a different duration because these syllables may be pronounced at different rates. Consequently, a straightforward method that compares the value of the first syllable at time t to that of the second bird syllable at the same relative time may not correctly classify a given syllable. Instead, an algorithm must be used to search the space of mappings from the time sequence of the first bird syllable to that of the second bird syllable. Dynamic time warping is a standard technique used to perform time alignment of two syllables with different duration (Vintsyuk, 1971).

In order for the DTW technique to be tested for bird syllable classification, a reference template must first be chosen and stored for every bird species. The classification process entails matching the incoming bird syllable with the stored templates. The templates and the incoming bird syllable are represented as a sequence of parameter vectors and the best matching template is the one that exhibits the minimum path aligning the input bird syllable to the template. The search space for the DTW method is constrained in such a way that the mapping function between the time axis of the input signal and the time axis of the template must be monotonically increasing with time so that the ordering of events in both the input and the template are preserved. The global distance score for a mapping path is simply the sum of local distances that make up the path.

The first preprocessing step in using DTW is the creation of bird syllable templates to be identified from the input bird syllable. In the current study, templates were selected manually to achieve the best recognition results. The second preprocessing step is to extract feature vectors for the templates and the input bird syllable. In this step, a digital recording of the syllable with 16-bit samples and a 16 kHz sample rate is first high pass filtered with a 100 Hz cutoff in order to remove the low frequency background noise attributable to wind or mechanical sounds. The filtered signal is then segmented into Hamming windowed 320-sample (20 ms) frames with 160-sample overlap (10 ms frame hop). The energy-based detection method mentioned in Sec. IV is used to detect the onset and offset of the sound signal. The frames between the onset and offset are transformed into the feature vectors on a frame-by-frame basis. Thirteen LPC coefficients $[a_1 \, a_2 \ldots a_P]$ $(P=13)$ are calculated using Durbin's recursive algorithm for every frame. To improve the recognition accuracy, the LPC coefficients are transformed into 18 LPCC coefficients $[c_1 \, c_2 \ldots c_P \ldots c_L]$ $(L=18)$ using the following equations:

$$c_1 = a_1, \tag{3}$$

$$c_n = a_n + \sum_{k=1}^{n-1} \left[\frac{n-k}{n} \cdot a_k \cdot c_{(n-k)} \right], \quad 2 \le n \le P, \tag{4}$$

$$c_n = \sum_{k=1}^{P} \left[\frac{n-k}{n} \cdot a_k \cdot c_{(n-k)} \right], \quad P+1 \le n \le L. \tag{5}$$

Note that the first order LPCC coefficient represents the spectral energy, which is generally not normalized between input signals, so c_1 is not used for the distance computation described next.

After generating the feature vectors for the templates and input bird syllables, the comparison is made by nonuniformly adjusting the time axis of the input syllable to achieve the best match to the template. In the matching process for an input bird syllable with M frames and a template with N frames, the time frames of the input bird syllable and the time frames of the templates are organized in a lattice (i,j), where i and j are the indexes of the time frames of the input syllable and the template, respectively. The quality of the match is measured recursively by the formula

$$D(i,j) = \min[D(i-1,j-1), D(i-1,j), D(i,j-1)]$$
$$+ d(i,j), \tag{6}$$

where $d(i,j)$ is the Euclidean distance between the two multidimensional vectors of the input signal at time frame i and the template at time frame j, and therefore $D(i,j)$ is the global distance up to (i,j). Given the initial condition $D(1,1) = d(1,1)$, we have the basis for an efficient recursive algorithm for computing $D(i,j)$. The final global distance $D(M,N)$ gives us the overall matching score of the template with the input. The input syllable is then classified as the species corresponding to the template with the lowest matching score.

2. Description of the classification system based on MFCC and HMM

The Hidden Markov Model is a doubly stochastic process with an underlying "hidden" stochastic process that is not observable, but can only be observed through another set of stochastic processes that produce the sequence of observed symbols (Rabiner, 1989). The HMM comprises a finite set of states, each of which is associated with a specific probability distribution. Transitions among the states are governed by a set of transition probabilities. In a particular state an observation can be generated according to the associated probability distribution.

The HMM can be used to model the bird sound generation statistically. It has been used for bird sound recognition due to its ability to characterize bird sounds in a mathematically tractable manner (Kogan and Margoliash, 1998). In contrast to the deterministic template matching of DTW, HMM uses a statistical representation. Therefore, this model can accumulate more information and possibly generalize better than techniques based on fixed templates.

The HMM procedure implemented for this study includes the same preprocessing as for the DTW procedure: 16-bit samples, 16 kHz sample rate, and 100 Hz high pass filtering, and the energy-based onset/offset detection described in Sec. IV. The reference templates are selected manually and include three sound files from each Category as the training cadre. Each preprocessed signal is segmented

TABLE III. Performance of a classification system based on linear prediction cepstral coefficients (LPCC) and dynamic time warping (DTW) methods (23/240 means that the error count was 23 out of 240).

(SNR) (dB)	12 natural sounds		28 total sounds	
	Noisy training set	Clean training set	Noisy training set	Clean training set
Original	23/240	23/240	33/560	33/560
30	23/240	24/240	33/560	36/560
24	27/240	28/240	38/560	40/560
18	30/240	34/240	43/560	48/560
12	38/240	40/240	54/560	57/560
9	50/240	61/240	68/560	82/560
6	59/240	90/240	81/560	116/560
3	68/240	115/240	93/560	148/560

TABLE IV. Performance of a classification system based on the mel-frequency cepstral coefficients (MFCC) and hidden Markov models (HMM) (12/240 means that the error count was 12 out of 240).

SNR (dB)	12 natural sounds		28 total sounds	
	Noisy training set	Clean training set	Noisy training set	Clean training set
Original	12/240	12/240	20/560	20/560
30	12/240	12/240	20/560	21/560
24	15/240	16/240	23/560	25/560
18	17/240	19/240	27/560	30/560
12	23/240	28/240	33/560	39/560
9	34/240	51/240	45/560	65/560
6	45/240	83/240	62/560	99/560
3	58/240	126/240	87/560	158/560

into Hamming windowed 256-sample (16 ms) frames with 128-sample overlap (8 ms frame hop). The frames between the onset and offset are transformed into 13-dimensional mel-frequency cepstral coefficient feature vectors using 26 mel filterbanks, on a frame-by-frame basis. The MFCC coefficients are a compact representation, which is the result of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale (Zheng *et al.*, 2001). As with the first order LPCC coefficient in the DTW method, the first order MFCC coefficient is not considered in the distance computation.

A continuous HMM representation was used in the traditional HMM-based classification system employed for this study. The HMM was a left-to-right model with six states. The densities of the observation probabilities in the emitting states were modeled as mixtures of two multidimensional Gaussian distributions with diagonal covariance matrices. The Baum-Welch algorithm was used in the training step employing the three manually selected bird syllables. One set of HMM parameters were generated for every bird species. The trained HMM parameters were then used for recognition using the Viterbi algorithm (Viterbi, 1967).

3. Simulation tests using the conventional classifiers

The sensitivity of the classification algorithms to noise was evaluated with an additional group of simulations. For each simulation the 560 example files (including the Template files) were deliberately contaminated with separate segments of uniformly distributed (white) noise to achieve a specified signal-to-noise ratio (SNR). The simulations included SNRs from 30 to 3 dB in steps of 3 or 6 dB.

A more likely practical situation would occur when using a set of "clean" Templates but testing with noisy Category recordings, i.e., field recordings that would ordinarily be obtained under less than ideal conditions. Thus, a separate simulation was used to examine this condition.

The DTW performance on the bird sound database is shown in Table III. The results indicate a likelihood of error for DTW with the original recorded signals of approximately 1/10 (i.e., 90% correct matches), with performance degrading to 71% correct classification under low signal-to-noise conditions.

The performance on the sound database for the conventional HMM system is shown in Table IV. The results show a likelihood of error for HMM of approximately 1/20 (95% correct matches) under high signal-to-noise conditions, degrading to 76% correct with noisy data.

Although the conventional procedures are moderately successful on the sound database used in this study and correspond well to prior results reported in the literature, the performance degradation with decreasing SNR is a serious shortcoming for a practical, robust system.

C. Classification by the spectral peak track method

The spectral peak track method requires a reference template for each bird species in the database. One representative syllable from each of the 28 Categories was selected manually, designated Template 0 through Template 27, and used as the training cadre.

The first test of the peak track procedures described in Sec. IV was to distinguish between single part and two-part syllables. The simple energy based detection algorithm performs well with the bird sounds in the database: all 560 example sounds were correctly classified. However, only 40 examples (Category 11 and Category 27) contained two-part syllables, so additional verification will be needed before assuming the simple algorithm is sufficient.

In the classification experiment, each of the 560 syllable recordings was processed using the peak track method described above. Here 238 out of 240 natural bird sounds were correctly classified (error rate 2/240), while the error rate was 5/560 (99% correct matches) for the entire natural and synthetic database at a high SNR. These results indicate that the peak track method was well suited to this particular database.

The five misclassified examples were examined. The first misclassification was a Category 3 example mistaken for Category 0. The particular example contained a frequency change from 3000 to 2500 Hz, while the Template 3 reference had a 3000 to 2000 Hz change. The resulting distance was closer to Template 0, causing the mismatch.

The second error was due to an example from Category 0 being mistaken for Category 1. In this case there are two peak tracks in Template 0, while the processed input example was found to have three significant peak tracks. The differing

TABLE V. Performance of the spectral peak tracks method (2/240 means that the error count was 2 out of 240).

SNR (dB)	12 natural sounds		28 total sounds	
	Noisy training set	Clean training set	Noisy training set	Clean training set
Original	2/240	2/240	5/560	5/560
30	3/240	3/240	6/560	6/560
24	5/240	5/240	8/560	8/560
18	4/240	4/240	7/560	7/560
12	3/240	3/240	6/560	6/560
9	6/240	6/240	9/560	9/560
6	8/240	14/240	9/560	14/560
3	11/240	18/240	11/560	18/560

number of peak tracks caused this example to have a greater distance from Template 0 than from Template 1 that has three peak tracks, therefore causing the misclassification.

The other errors were because of examples from Category 11 being mistaken for Category 27. The Category 11 examples are a natural bird sounds, while Category 27 contains synthesized examples with relatively strong peak track powers that were included specifically as distracters (successful in triggering a misclassification in this case). The Template 11 data contains two peak tracks, while the processed input examples were found to have four significant peak tracks. The differing number of peak tracks caused these examples to have a greater distance from Template 11 than from Template 27, thereby causing the misclassifications.

The classification performance degraded as the SNR decreased from 30 to 24 dB. When the SNR decreased from 24 to 12 dB, the classification accuracy unexpectedly improved, despite the degraded signal quality. An examination of the Category 11 signal revealed that the additive noise slightly altered the relative power of the peak tracks, causing the distance between the Category 11 signal and Template 11 to be less than the distance between Category 11 and Template 27. A similar subtle change occurred between a Category 3 signal and Template 0, causing a correct classification. Decreasing the SNR to 9, 6, and 3 dB degraded the classification accuracy, although the overall performance was still good (95% correct at 3 dB SNR), as shown in Table V.

When using "clean" Templates but testing with noisy Category recordings, the classification accuracy was unaffected until the SNR decreased to 6 and 3 dB. The performance differed between the noisy and clean Template cases, but not in a significant manner. Thus, based on this simulated noise test the peak track method does not appear to be particularly sensitive to the SNR difference between the Template and the test samples.

D. Discussion

As indicated by the results in Tables III, IV, and V, the spectral peak track method provided better overall results than the conventional DTW and HMM methods, particularly in the low signal-to-noise ratio tests. There are two expected reasons for this result. First, the conventional methods based on a linear prediction model have difficulty with the sparse spectrum of the bird syllables in the test database: insufficient information is present in each syllable recording to create a unique and easily distinguishable LPC model. The frequency spacing of the spectral components is relatively wide compared to the underlying resonances of the bird vocal tract, indicating that the linear prediction methods are mismatched to the signals in the database. Second, the conventional model parameters are quite sensitive to existing background noise, reverberation, and competing sounds in the recordings. The results in Tables III and IV for the 12 natural sounds and for the 28 total sounds show that the conventional systems operated better on the artificial distracter signals than on the natural sounds with inherent background noise. For these reasons it is useful to consider alternative pattern classification techniques, such as the spectral peak track method presented in this paper.

VI. CONCLUSION

The spectral peak track method described in this paper appears to work as designed for isolated bird syllables, and the simulation results are better in comparison to conventional DTW and HMM methods used to classify the same database. The proposed method extends the prior peak track methods by using a variable number of tracks to represent each syllable by determining the relative power of each detected spectral track. This method accommodates tonal, harmonic, or inharmonic syllables, and bases the pattern classification on the strongest tracks present in each syllable. The set of parameters derived for each significant peak track was quite robust in the presence of simulated additive noise, which is an encouraging result for future applications of this technique in classifying bird vocalizations.

However, the proposed method is inappropriate for use with bird vocalizations containing aperiodic or noise-like components because the assumption of connected peak tracks is violated in these cases. The proposed method is also inappropriate if the underlying spectral components change too rapidly in frequency or fluctuate in amplitude such that the peak tracks cannot be determined reliably. Research is continuing on methods to classify such rapidly varying signals, and also to identify broadband and noisy sounds.

In the current system we manually extract one syllable from the recorded bird sounds and save the data in a separate sound file. However, in an on-line real time production application it will be necessary to demonstrate an automatic syllable extraction method. Thus, there remain a variety of engineering challenges to deploying this system for real-time classification.

Anderson, S. E., Dave, A. S., and Margoliash, D. (1996). "Template-based automatic recognition of birdsong syllables from continuous recordings," J. Acoust. Soc. Am. 100, 1209–1219.

Catchpole, C. K., and Slater, P. J. B. (**1995**). *Bird Song: Biological Themes and Variation* (Cambridge University Press, Cambridge, UK).

Duda, R. O., Hart, P. E., and Stork, D. G. (**2001**). *Pattern Classification*, 2nd ed. (Wiley, New York).

Ellis, D. P. W. (**2003**). "Sinewave and sinusoid+noise analysis/synthesis in MATLAB," Electronic document, URL: ⟨http://www.ee.columbia.edu/~dpwe/resources/matlab/sinemodel⟩.

Fagerlund, S. (**2004**). "Automatic recognition of bird species by their sounds," Masters thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki Univ. of Technology, Laboratory of Acoustics and Audio Signal Processing.

Härmä, A. (**2003**). "Automatic identification of bird species based on sinusoidal modeling of a syllable," *IEEE Int. Conf. Acoust. Speech and Signal Processing (ICASSP 2003)*, **5**, 545–548.

Härmä, A., and Somervuo, P. (**2004**). "Classification of the harmonic structure in bird vocalization," *IEEE Int. Conf. Acoust. Speech, Signal Processing (ICASSP 2004)*, **5**, 701–704.

Ito, K., Mori, K., and Iwasaki, S. (**1996**). "Application of dynamic programming matching to classification of budgerigar contact calls," J. Acoust. Soc. Am. **100**, 3947–3956.

Kahrs, M., and Avanzini, F. (**2001**). "Computer synthesis of bird songs and calls," *Proc. Conf. Digital Audio Effects (DAFx-01)*, pp. 23–27.

Kogan, J. A., and Margoliash, D. (**1998**). "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," J. Acoust. Soc. Am. **103**, 2185–2196.

Krebs, J. R., and Kroodsma, D. E. (**1980**). "Repertoires and geographical variation in bird song," Journal of Advances in the Study of Behavior **11**, 143–177.

Kroodsma, D. E., and Miller, E. H. (**1996**). *Ecology and Evolution of Acoustic Communication in Birds* (Comstock, Ithaca, NY).

McAulay, R. J., and Quatieri, T. F. (**1986**). "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Process. **34**, 744–754.

McIlraith, A. L., and Card, H. C. (**1997**). "Birdsong recognition using back-propagation and multivariate statistics," IEEE Trans. Signal Process. **45**, 2740–2748.

National Wind Coordinating Committee (NWCC) (**2004**). "Wind-turbine interactions with birds and bats: a summary of research results and remaining questions," RESOLVE, Washington, DC.

Nowicki, S. (**1997**). "Bird acoustics," in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York), Chap. 150, pp. 1813–1817.

Pascarelle, S. M., Pinezich, J., Merritt, R. L., Kelly, T. A., Roman, B., and Maher, R. C. (**2004**). "Automated acoustic monitoring of bird strike hazards," *6th Annual Meeting of the Bird Strike Committee USA/Canada*, Baltimore, MD, September, **2004**.

Rabiner, L. R., Rosenberg, A. E., and Levinson, S. E. (**1978**). "Considerations in dynamic time warping algorithms for discrete word recognition," IEEE Trans. Acoust., Speech, Signal Process. **26**, 575–582.

Rabiner, L. R. (**1989**). "A tutorial on hidden markov models and selected applications in speech recognition," Proc. IEEE **77**, 257–286.

Rabiner, L. R., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice–Hall, Englewood Cliffs, NJ).

Smith, J. O., and Serra, X. (**1987**). "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," *Proc. Int. Computer Music Conf.*, San Francisco, Computer Music Association.

Vintsyuk, T. K. (**1971**). "Element-wise recognition of continuous speech composed of words from a specified dictionary," Journal of Cybernetics and Systems Analysis, **7**(2), 361–372.

Viterbi, A. J. (**1967**). "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," IEEE Trans. Inf. Theory **13**, 260–269.

Zheng, F., Zhang, G. L., and Song, Z. J. (**2001**). "Comparison of different implementations of MFCC," Journal of Computer Science & Technology, **16**(6), 582–589.

Erratum: "Atmospheric absorption in the atmosphere up to 160 km" [J. Acoust. Soc. Am. 115(3), 1012–1032 (2004)]

Louis C. Sutherland
Consultant in Acoustics, 27803 Longhill Drive, Rancho Palos Verdes, California 90275-3908

Henry E. Bass
National Center for Physical Acoustics, The University of Mississippi, University, Mississippi 38677

Equation (8) should read

$$\beta_1/\beta_0 = \{1/2[(1 + \nu^2)^{1/2} + 1]/(1 + \nu^2)\}^{1/2}. \tag{8}$$

Equation (16) should read

$$\alpha_{cl} = \alpha_1(\beta_2/\beta_0) = (2\pi f/c_0)\left[\frac{1/2[(1 + \nu^2)^{1/2} - 1][1 + (x')^2]}{[1 + \nu^2][1 + (\sigma x')^2]}\right]^{1/2}. \tag{16}$$

The first sentence in the paragraph on page 1014 that begins "The second term in Eq. (14) is the attenuation coefficient. . ." should read "The second term in Eq. (15) is the attenuation coefficient. . .".

Equation (18) should read

$$\alpha_{rot} = \alpha_2(\beta_1/\beta_0) = (2\pi f/c_0)X_{ON}\left[\frac{(\sigma^2 - 1)x}{2\sigma}\right]\left[\frac{1/2[(1 + \nu^2)^{1/2} + 1]}{[1 + \nu^2][1 + (\sigma x')^2][1 + (x')^2]}\right]^{1/2}. \tag{18}$$

In Section IV-A, on page 1017, the line just above Eq. (27) should end with a reference: ". . .and the atmospheric temperature T, as[7]. . .".

In the paragraph below Eq. (27) on page 1017, for consistency, the symbol for the normalized relaxing specific heat on line 4 should be given as C_i'/R and similarly on lines 3 and 4 in the next paragraph and in the caption for Figure 3.

In Section IV-B4 on page 1019, after Eq. (31), units should be added to L so that the line should read ". . .where $L = \exp(-7.72T_r)$, Hz."

In Table II, the entry under column A_0 for Pressure, Leg V should be 7.63380E-01.

These corrections are all typographical in nature and do not affect any of the figures, which were plotted using the correct equations shown above.

Session 1aID

Interdisciplinary: Opening Ceremonies, Plenary Lectures

Whitlow W. L. Au, Cochair

Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

Sadaoki Furui, Cochair

Tokyo Inst. of Technology, Dept. of Computer Science, 2-12-2 O-akayama Meguro-ku, Tokyo 152-8552, Japan

Opening Ceremony—8:00

Chair's Introduction—8:10

Invited Paper

8:15

1aID1. How can auditory presence be generated and controlled? Masayuki Morimoto (Environ. Acoust. Lab., Faculty of Eng., Kobe Univ., Nada, Kobe 657-8501 Japan)

This paper reviews key results of many listening tests on auditory presence, especially auditory localization (AL) and auditory spatial impression (ASI) by the author. The author gave the first demonstration that AL in any direction can be simulated through two loudspeakers using head-related transfer functions (HRTFs). However, individual differences in HRTFs affect the accuracy of AL. It is basically possible to localize sound images in any direction using median-plane HRTFs combined with interaural differences. Furthermore, HRTFs can be simplified by the combination of only some spectral peaks and notches. Meanwhile, the author demonstrated that ASI comprises at least two perceptual components. One is auditory source width (ASW), defined as the width of a sound image fused temporally and spatially with the direct sound image. The other is listener envelopment (LEV), which is the degree of fullness of sound images around the listener, excluding the sound image composing ASW. A listener can perceive these two components separately. The perception of ASW and LEV has a close connection with the law of the first wavefront. Acoustic components under the upper limit of the law contribute to ASW, and acoustic components beyond the upper limit of the law contribute to LEV.

Chair's Introduction—9:05

Invited Paper

9:10

1aID2. Therapeutic ultrasound. Lawrence Crum (Ctr. for Industrial and Med. Ultrasound, Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105, lac@apl.washington.edu)

The use of ultrasound in medicine is now quite commonplace, especially with the recent introduction of small, portable, and relatively inexpensive, hand-held diagnostic imaging devices. Moreover, ultrasound has expanded beyond the imaging realm, with methods and applications extending to novel therapeutic and surgical uses. Among these applications are tissue ablation, acoustocautery, lipoplasty, site-specific and ultrasound mediated drug activity, extracorporeal lithotripsy, and the enhancement of natural physiological functions such as wound healing and tissue regeneration. A particularly attractive aspect of this technology is that diagnostic and therapeutic systems can be combined to produce totally noninvasive, image-guided, bloodless surgery. This general lecture will review a number of these exciting new applications of ultrasound and address some of the basic scientific questions and future challenges in developing these methods and technologies for general use in our society. We shall particularly emphasize the use of high-intensity focused ultrasound (HIFU) in the treatment of benign and malignant tumors as well as the induction of acoustic hemostasis, especially in organs that are difficult to treat using conventional medical and surgical techniques. [Work supported in part by the NIH, NSBRI, ONR, and DARPA.]

Session 1aAA

Architectural Acoustics: Computer Modeling of Room Acoustics I

Lily M. Wang, Cochair

Univ. of Nebraska Lincoln, 200B Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681

Shinichi Sakamoto, Cochair

Univ. of Tokyo, Inst. of Industrial Science, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Chair's Introduction—10:55

Invited Papers

11:00

1aAA1. A transparency model and its applications for simulation of reflector arrays and sound transmission. Claus Lynge Christensen and Jens Holger Rindel (Odeon A/S, Oersted-DTU, Bldg. 352, 2800 Lyngby, Denmark)

The paper describes a new method for simulating the frequency-dependent reflection and transmission of reflector arrays, and the frequency-dependent airborne sound insulation between rooms by means of a room acoustic computer model. The method makes use of a transparency method in the ray-tracing process. In the first step of the calculation the rays hitting the relevant surfaces may either be reflected or transmitted, using a probability of 50%. In the next step the impulse responses in the receiver positions are calculated using a frequency-dependent correction to account for the reflected or transmitted energy. The method applied for the reflector array is based on a theoretical model that takes into account the dimensions of the reflecting surface, path lengths, and angle of incidence. The transmission calculation is based on the users' data for the frequency-dependent transmission loss of the partition, and this is useful for the auralization of sound transmission through different building constructions. The acoustic properties like volume, reverberation time, and the area of the transmitting surfaces are included in the simulation.

11:20

1aAA2. Numerical determination of scattering coefficients of wall surfaces for geometrical room acoustic simulation. Tetsuya Sakuma, Yoshiyuki Kosaka, and Yuki Tachioka (Grad. School of Frontier Sci., Univ. of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8563, Japan, sakuma@k.u-tokyo.ac.jp)

There exist a dozen geometrical room acoustic simulation programs which include scattering coefficients of wall surfaces to generate more realistic sound energy reflection. In the light of this utilization, a measurement method of the random-incidence scattering coefficient has been recently standardized by ISO 17497-1, while a numerical technique with BEM is developed to determine the scattering coefficient. First, one case study is done to investigate the behavior of scattering performance of periodical surfaces with sinusoid, triangles and rectangles, with changing the height of surface roughness. As a result, it is seen that the height-to-period ratio of about 30% maximizes the scattering coefficient in the range of middle and high frequency. Second, another case study of geometrical room acoustic simulation is done for a rectangular room composed of uneven reflective walls and absorptive ceiling and floor. Geometrical simulation is performed on a variety of conditions with changing the scattering coefficient, and the reverberation times simulated are compared with those given by wave-based simulation at every octave bands. From the correspondence in reverberation time, the scattering coefficients of the walls are estimated, and its frequency characteristics are compared with those given by the former numerical determination.

11:40

1aAA3. Edge diffraction in computer modeling of room acoustics. U. Peter Svensson (Acoust. Group, Dept. of Electron. and Telecommunications, Norwegian Univ. of Sci. and Technol., NO-7491 Trondheim, Norway, svensson@iet.ntnu.no) and Paul T. Calamia (Rensselaer Polytechnic Inst., Troy, NY)

Computer modeling in room acoustics is typically based on geometrical acoustics techniques. Limitations with such methods include, among other things, a lack of diffraction modeling, which primarily leads to inaccuracies at low frequencies. The inclusion of diffraction modeling is quite straightforward for first-order diffraction, which can be combined with specular reflections of any order. One impractical aspect, however, is that the number of diffraction components can be extremely high, and grows faster (with the reflection order) than the number of specular reflections does. At the same time, the importance, or magnitude, of the diffraction components will differ over an immense range. This variation can be exploited by estimating the importance of each diffraction contribution by the magnitude of its onset, and skipping the remainder of the calculations for those that are deemed too weak. This will be demonstrated for some typical geometrical cases including a set of overhead reflectors, an orchestra pit, and a convex hall shape for which diffraction is less critical. Results indicate that "diffraction culling" can provide a significant reduction in computation time with only small effects on the overall responses for the tested geometries. [This research has been supported by the Research Council of Norway].

Session 1aAB

Animal Bioacoustics and Underwater Acoustics: Remote Monitoring of Biological Sounds I

Marc O. Lammers, Chair

Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

Contributed Papers

10:30

1aAB1. Assessing marine mammal abundance, density and habitat use with acoustics. Robert Gisiner (Office of Naval Res., 875 North Randolph St., Ste. 1425, Arlington, VA 22203)

Counting marine mammals is challenging: when seen they usually cannot be heard, and when heard they usually cannot be seen. Both methods miss some fraction of the population, but how does one reconcile two sampling methods that are uncalibrated themselves and will be difficult to calibrate against each other? Three approaches to calibration are reviewed. The one-way calibration approach treats one method as correct and derives a correction function for estimating population data from the other method. The parallel independent approach derives sources of error for each method independently and then compares population estimates: close agreement would strengthen confidence in both methods. The concurrent approach attempts to reconcile each visual detection to an acoustic detection and vice versa, with estimated total numbers derived from the intersection of the two sets. Examples of each method will be provided to illustrate the challenges and potential of each approach.

10:45

1aAB2. Reducing source localization errors: A visualization method to help guide the design, calibration, and use of widely separated acoustic sensor arrays. Catherine L. Berchok, Gerald L. D'Spain, and John A. Hildebrand (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 93940-0701, cberchok@ucsd.edu)

The use of acoustic arrays to obtain source positions of vocalizing animals has become quite common in field studies of wild animal populations, but the results presented rarely include error bounds, leading to a false sense of confidence in the animal locations. As research from the acoustics community has shown, these source localization errors vary with respect to bearing and distance from the array and are generated by measurement uncertainties associated with the propagation environment, field equipment, and data processing/analysis methods. However, these errors can be reduced by making informed decisions about the design, calibration, and deployment of the arrays and by including measurements of amplitude ratios and vector sensor bearings in addition to time-of-arrival differences. A numerical method is presented here that allows for visualization and quantification of the impact of measurement uncertainties on source localization error bounds for widely spaced acoustic sensor array. This method will be used to show how quick comparisons of source localization error bound maps can help determine appropriate array designs for a particular field application. Among the parameters compared are array geometry, beamforming techniques, and measurement of various combinations of properties of the received acoustic field. [Work supported by a Hunt Postdoctoral Fellowship.]

11:00

1aAB3. Detection distances of the sounds of large whales recorded on hydrophones in the offshore Gulf of Alaska. Kathleen Stafford, Sue Moore (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98115, stafford@apl.washington.edu), and David Mellinger (Oregon State Univ., Newport, OR 97365)

The current status of most species of endangered baleen whales, including blue, fin, and humpback whales, in the Gulf of Alaska is unknown due to a lack of basic information on distribution and seasonal abundance. Remote passive acoustic monitoring can provide this information for vocal whales. However, to begin to estimate an index of abundance of calling animals, the distance at which they can be detected needs to be determined. In order to estimate transmission loss, a parabolic equation acoustic propagation model was used to provide mean loss estimates along four transects at $0.1a$ intervals to $5a$ to the N, E, S, and W from each of four receivers moored in the sound channel for species-specific frequencies and at depths at which the animals are thought to produce sound. For all species and locations, the detection range was largely determined by the choice of ambient noise levels. This suggests that masking due to anthropogenic noise could limit the range over which these animals can be detected by the moored instrument and, more importantly, the range over which they might communicate.

11:15

1aAB4. An archive of odontocete sounds for automatic call recognition. David K. Mellinger, Sharon L. Nieukirk, and Sara L. Heimlich (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, david.mellinger@oregonstate.edu)

Acoustic remote monitoring using automatic detection of cetacean sounds is becoming widely used to study species' seasonal distributions and to monitor and mitigate human impacts on cetaceans. To date, most work on automatic detection has focused on baleen whales and a handful of odontocetes, principally sperm whales and harbor porpoises. To further the development of detection methods for other odontocetes (particularly beaked whales, which have become the focus of much interest), it would be helpful to have standardized training and testing datasets. In preparing such datasets, several questions must be addressed: What recordings are available? How certain are the species IDs for these, and how should differing levels of certainty be handled? What is the recording quality? How should recordings be presented and annotated to make them most useful for automatic detection research? Are there any recordings known to be made on-axis with respect to the animal? How are multiple overlapping animals handled? How are mixed-species groups handled? Are data needs different for clicks and whistles? What metadata will be most useful for detection? For localization? How should recordings and metadata be stored and made them available on the net? Partial or full solutions to these questions will be presented.

1aAB5. Automatic vocal individual recognition of acorn woodpecker (*Melanerpes formicivorus*) based on hidden Markov models. Yuan Yao (Dept. of Ecology and Evolutionary Biol., Univ. of California, Los Angeles, Los Angeles, CA 90095, yaoyuan@ucla.edu), Ying Lin (Univ. of Arizona, Tucson, AZ 85721), Andreas Ali, and Charles Taylor (Univ. of California, Los Angeles, Los Angeles, CA 90095)

The acorn woodpecker (*Melanerpes formicivorus*) is a highly social cooperatively breeding species. A variety of conspecific interactions inside and between family groups is mediated by vocalizations. In the current study, acoustic sensor networks are used to monitor vocal behaviors of acorn woodpecker. In order to identify the callers in our behavioral study, we present a method for the automatic vocal individual recognition using hidden Markov models (HMMs). Field recordings from eight woodpeckers in two family groups were made in northern California. Individual identification accuracy is 92.65% for isolated waka calls, 84.31% for isolated syllable wa, 73.04% for isolated syllable ka, using 7-state HMMs with MFCC_E_D parametrization. For the continuous recordings, the recognition rate is 66.67%–100.00%. We studied the influence of two major factors in the performance of the HMMs: (i) different structures of HMMs and parametrizations of data, (ii) different qualities of signals, including degraded calls after propagation and calls from two simultaneous sources split by using beamforming of sensor networks. These results suggest that HMMs, combined with sensor networks, are a promising tool for data collection in vocal behavioral studies of animals.

1aAB6. Cetacean vocalization tracking using an adaptive Poisson rate estimator. Owen P. Kenny and Craig R. McPherson (Dept. of Elec. and Comput. Eng., James Cook Univ., Douglas 4811, Queensland, Australia)

This paper presents algorithms for tracking the formant structure of cetacean vocalizations using auditory modeling and adaptive filtering to estimate the Poisson intensity of nerve firing rates. These rates are in turn related to the instantaneous frequency of the formant vocalisation. The auditory model considered in this paper produces a process described as a spike train representing the auditory nerve firing, the Poisson intensity of which is determined by an underlying modulation process. Previous work [Kenny *et al.*, *Adaptive Filter for Speech Enhancement Using Poisson Rates from an Auditory Model*, paper presented at the WESPAC VIII, Melbourne, Australia (2003); C. McPherson, *Development of a Toolkit for Analysing and Localising Underwater Acoustic Signals* (James Cook University, Townsville, 2005)] has shown that this modulation process can be evaluated from the conditional expectation of the Poisson rate given inter-spike arrival time observations. The paper also addresses the application of adaptive filters to update the estimator parameters for different acoustic environments, which is unknown. These algorithms have been applied to real field recordings of interacting false killer whales (*Pseudorca crassidens*) and results from experiment of shown individualisation and extraction of individual narrow-band vocalizations.

TUESDAY MORNING, 28 NOVEMBER 2006   MAUI ROOM, 11:00 A.M. TO 12:05 P.M.

Session 1aAOa

Acoustical Oceanography and Underwater Acoustics: Acoustic Scattering by Aquatic Organisms I

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Masahiko Furusawa, Cochair
Tokyo Univ. of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo, 108-8477, Japan

Chair's Introduction—11:00

Invited Papers

11:05

1aAOa1. A broadband acoustic system for resonance classification of swimbladder-bearing fish. Timothy K. Stanton, Dezhang Chu (Appl. Ocean Phys. & Eng. Dept., Woods Hole Oceanogr. Inst., M.S. #11, Woods Hole, MA 02543, tstanton@whoi.edu), J. Michael Jech (Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543), and James D. Irish (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

A commercial system originally designed for surveying the seafloor has been adapted for measuring the swimbladder resonance of swimbladder-bearing fish. The new towed system has three broadband active acoustic echosounders with nearly continuous coverage of the frequency band 2–110 kHz (with some gaps). The lower frequency bands are used for measuring the resonance, while the highest band can detect zooplankton. A pulse-compression approach (cross correlating the transmit waveform with the echoes) is used to exploit the bandwidth through significantly improving the range resolution to 20, 10, and 2 cm (low-, med-, and high-frequency bands, respectively). The system has been successfully deployed as deeply as 200 m and observed resonances at approximately 3 kHz for herring. Because of the 10–20-cm range resolution at the low frequencies, fish near the towbody could be resolved from one another and fish swimming within 1 m of the seafloor could be distinguished from the seafloor. This system is compared with previous approaches of measuring swimbladder resonance. The potential of the resonance classification method for significantly reducing ambiguities associated with interpreting acoustic echo data compared with traditional high-frequency narrow-band systems (10's to 100's of kHz) is discussed.

11:25

1aAOa2. Effects of scattering layer composition, animal size, and numerical density on the frequency dependence of volume backscatter. Kelly Benoit-Bird (College of Oceanic and Atmospheric Sci., Oregon State Univ., Corvallis, OR 97331)

The mesopelagic boundary community around the Hawaiian Islands is a land-associated, sound scattering layer that undergoes diel migrations with both a vertical and horizontal component. A video camera system was developed to quantitatively examine the numerical density, size, and taxonomic composition of micronekton. The camera system was combined with a four-frequency vessel-mounted echosounder system (38, 70, 120, and 200 kHz) to document the full migration range of micronekton and describe the changes in composition and density throughout their diel migration. Migrating animals split into multiple, distinct layers at night with differences in micronekton density, composition, and size. These differences were correlated with differences in the frequency response of volume backscatter. The relationship between these variables and backscatter intensity relationship between frequencies is complex. The results suggest strong partitioning of habitat by these animals in space and time, which, along with the high densities of animals, indicate that competition is important in driving the behavior and structuring the community.

11:45

1aAOa3. Probing the deep: Acoustic characterization of the Mid-Atlantic Ridge ecosystem. John K. Horne, Cairistiona I. H. Anderson (School of Aquatic and Fishery Sci., Univ. of Washington, Box 355020, Seattle, Wa 98195), Olav Rune Godoe, and Ruben Patel (Inst. of Marine Res., Nordnes 5817, Bergen, Norway)

Implementation of the ecosystem approach to resource management potentially increases the use of acoustics during assessment surveys. A Census of Marine Life survey on the mid-Atlantic ridge during summer 2004 was an opportunity to develop a practical approach. The Norwegian vessel, *G. O. Sars*, is an acoustically-quieted platform equipped with a five-frequency (18, 38, 70, 120, and 200 kHz) echosounder, acoustic Doppler current profilers (ADCPs), multibeam sonars, and a deep-towbody. Paucity of information on species composition limited the ability to conduct a traditional acoustic survey. Acoustic structure independent of biological sampling was quantified and then integrated acoustic density and target strength observations with biological community composition and length frequency data. Species were not assigned backscatter thresholds or water column regions at the onset of analysis. Near-real-time products monitored biomass distributions: daily echograms, target strengths, and frequency-differenced echograms. Persistent biological layers at different depths occurred in the water column. Layers or components migrated toward the surface during dark hours. Modes of target strength frequency distributions and trawl catch compositions differed among layers. The use of acoustic and net technologies can be integrated with near-real-time analytic results to quantitatively characterize pelagic ecosystems. [Work supported by ONR, IMR, and NOAA Fisheries.]

TUESDAY MORNING, 28 NOVEMBER 2006　　　　　　　　　　KAUAI ROOM, 11:00 A.M. TO 12:05 P.M.

Session 1aAOb

Acoustical Oceanography and Signal Processing in Acoustics: Acoustic Tomography for Coastal and Deep Water Applications I

James A. Mercer, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Arata Kaneko, Cochair
Hiroshima Univ., Graduate School of Engineering,1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8527, Japan

Chair's Introduction—11:00

Invited Papers

11:05

1aAOb1. Current structure measurements by the coastal acoustic tomography. Arata Kaneko, Keisuke Yamaguchi, Ju Lin, Noriaki Gohda (Grad. School of Eng., Hiroshima Univ., 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan), Hong Zheng (SEA Corp., Ichikawa, Chiba 272-0127, Japan), and Yoshio Takasugi (Chugoku Ctr. of Natl. Inst. of Adv. Industrial Sci. and Technol., Kure, Hiroshima 737-0197, Japan)

The coastal acoustic tomography (CAT), coastal-sea application of deep-sea acoustic tomography (Munk and Wunsch, 1978), is proposed as an advanced technology to map current structures in the coastal sea. A sequence of successful experiments was recently carried out in the coastal seas with various oceanographic conditions around Japan (Yamaguchi *et al.*, 2005; Kaneko *et al.*, 2005; Lin *et al.*, 2005). In the Tokyo Bay experiment of November 2002, the 2-D tidal currents inducing a clockwise residual circulation were mapped by eight CAT systems, located at both the eastern and western coasts of the bay. The vortex-embedded tidal currents with a maximum velocity of 5 ms^{-1} were targets in the Kanmon Strait experiment of March 2003. In the Hiroshima Bay experiment of

September 2003, the stratified sea was shifted to the homogenized sea after the passage of a typhoon. The transition of tidal current structures before and after the typhoon was well measured by seven CAT systems, located at the northern coast and on the southern oyster rafts. The procedure of 2-D current mapping for the coastal seas with depths shallower than about 30 m is well established through a sequence of CAT experiments mentioned above.

11:25

1aAOb2. Shallow-water tomography: Is it a good idea? James F. Lynch and Timothy F. Duda (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, Jlynch@whoi.edu)

There seems to be a feeling among ocean acousticians that anything that can be measured acoustically should be measured acoustically. However, there are two drawbacks to that point of view, comforting as it might be to us acousticians. First, many things that one would think are measurable with acoustics are actually not measurable, or are only barely so. And second, acoustics may turn out to be a poor or cost-ineffective measurement compared to other techniques. In this paper, we will look at whether acoustic tomography is viable, can produce high-quality measurements, and is cost effective in shallow water. We will use feature resolution, cost, reliability, intrusiveness, averaging properties, and other measures to obtain a weighted description of how effective tomography is for typical shallow-water (coastal) applications and then compare this to the same measure for other (conventional) oceanographic measurement systems. [Work supported by ONR.]

11:45

1aAOb3. Ocean acoustic tomography using high-frequency systems. Michael B. Porter, Katherine H. Kim, Martin Siderius (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130, mikeporter@hlsresearch.com), Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), Thomas Folegot (NATO Undersea Res. Ctr., 19138 La Spezia, Italy), and the KauaiEx Group (NUWC)

New system concepts have emerged in which large numbers of acoustic and environmental sensors are networked together in littoral areas. The spatial scales for these networks are much finer—features such as linear and nonlinear internal waves/tides can affect both network and sensor performance. A vision readily comes together in which ocean acoustic tomography is used to monitor the finescale, using the network to relay travel-time information in real time and also to probe the environment. Further, travel-time information may be assimilated into an ocean circulation model together with direct current and temperature measurements from fixed and mobile platforms to provide nowcasts and forecasts of the local ocean weather. The acoustic tomography in this vision uses a much higher frequency band (e.g., 8–16 kHz) than has traditionally been used and there has been some skepticism about the feasibility of doing tomography in this band. Using data from two large experiments in Hawaii (KauaiEx and MakaiEx), we show that (1) HF signals can be used to clearly measure the channel impulse response, (2) changes in the impulse response are directly linked to changes in the oceanography, and, therefore, (3) HF signals can be used to invert for the ocean structure.

TUESDAY MORNING, 28 NOVEMBER 2006   KAHUKU ROOM, 10:30 A.M. TO 12:30 P.M.

Session 1aBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Enhancement of Drug Activity

Constantin C. Coussios, Cochair
Univ. of Oxford, Dept. of Engineering Science, Inst. of Biomedical Engineering, 43 Banbury Rd., Oxford, OX2 6PE, U.K.

Shin-ichiro Umemura, Cochair
Kyoto Univ., School of Health Science, Kyoto, 607-8507 Japan

Invited Papers

10:30

1aBB1. Ultrasound-enhanced thrombolysis. Christy K. Holland, Saurabh Datta, Sampada S. Vaidya (Dept. of Biomed. Eng., MSB 6167, 231 Albert Sabin Way, Cincinnati, OH 45208), Jason M. Meunier, George J. Shaw (Dept. of Emergency Medicine), and Constantin-C. Coussios (Univ. of Oxford, Oxford OX1 3PJ, UK)

Substantial enhancement of recombinant tissue plasminogen activator (rt-PA) thrombolysis can be achieved with ultrasound, suggesting its use as an adjunctive treatment in thrombolytic therapy for stroke. This enhancement may be due to acoustic cavitation, thermal effects, forced convection of rt-PA, or biochemical enhancement. The effect of US duty cycle on the lytic efficacy of pulsed 120-kHz ultrasound in a human *in vitro* clot model was investigated, using a time-lapse microscopic imaging technique. The initial lytic rate increased linearly with duty cycle. In addition, a passive cavitation detection scheme was used to explore stable and inertial cavitation as potential mechanisms for ultrasound-enhanced thrombolysis. Porcine clots were exposed to pulsed ultrasound and rt-PA *in vitro* for 30 min using four separate ultrasound treatment regimes: (i) no cavitation (0.15 MPa); (ii) stable cavitation alone (0.24

MPa), (iii) stable and inertial cavitation combined (0.36 MPa); or (iv) no ultrasound. Percent clot mass loss after each treatment was used to determine thrombolysis efficacy. Clots exposed to stable cavitation and rt-PA exhibited the highest mass loss. Significant enhancement of thrombolysis correlates with presence of cavitation and stable cavitation appears to play a more important role in the enhancement of thrombolysis.

10:50

1aBB2. *In situ* microbubble generation from liquid precursor for diagnostic and therapeutic application. Ken-ichi Kawabata, Akiko Yoshizawa, Takashi Azuma, Hideki Yoshikawa, Hiroko Watanabe, Kazuaki Sasaki, Koji Hirata, and Shin-ichiro Umemura (Central Res. Lab., Hitachi, Ltd., Japan)

Stabilized microbubbles show high echogenicity and characteristic nonlinear acoustic responses and also work as "sensitizers" for HIFU therapy. Accordingly, if microbbles were selectively placed in targets inside the body, targeted diagnosis and therapy would be possible. However, unfortunately, microbubbles are too large to apply tissues other than in vessels. For tumor detection and therapy, we propose the use of a nano-sized liquid precursor of microbubbles, which is small enough to accumulate in tumor tissues and generate microbubbles with ultrasound pulses. We have developed emulsion-based precursors containing perfluoropentane and found that microbubbles can be produced with a 10 μ pulse of 2-MHz ultrasound at intensities of more than about 150 W/cm^2 in water. The intensity threshold was controllable by the added nonvolatile fluorocarbons. The microbubble generation was also observed in mice tumors at the same ultrasound intensity range. It was also found that when microbubbles were generated in tumor tissues, CW ultrasound exposure resulted in tumor damage. Our results indicate the potential usefulness of our precursor for targeted ultrasound diagnosis and therapy. [Part of this work was supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology through a grant-in-aid for the creation of innovations through business-academic public sector cooperation.]

11:10

1aBB3. Developing novel therapeutic strategies through biological effects of ultrasound. Sumit Paliwal and Samir Mitragotri (Dept. of Chemical Eng., Univ. of California, Santa Barbara, CA 93106)

Ultrasound-mediated cavitation has been shown to facilitate delivery of drugs across biological barriers in cells and tissues. However, these nonthermal mechanisms can also trigger various biochemical pathways in the biological milieu, opening up doors to interesting ultrasonic therapeutic applications. For example, a brief application of ultrasound followed by treatment with quercetin (a poorly potent chemotherapeutic drug) becomes selectively cytotoxic towards prostate and skin cancer cells. As opposed to the increased delivery of drugs to cancer cells, this treatment methodology is hypothesized to be mediated through inhibition of heat shock protein-hsp72, a vital stress-protein essential for the survival of cancer cells. Additional examples demonstrating the use of ultrasound's biological effects for therapeutic applications will be discussed. For example, ultrasound is demonstrated to work as a vaccination adjuvant by activating the Langerhans cells (immune cells present in skin) and leading to a robust immune response against tetanus toxoid. Overall, this presentation will demonstrate the increasing importance of ultrasound's role as a biological sensitizer enabling novel therapeutic strategies, a role which is beyond its conventional use as a drug delivery tool.

Contributed Papers

11:30

1aBB4. Ultrasound-mediated nanoparticle drug delivery in an *in vivo* tumor model. Balasundar I. Raju, Christopher S. Hall (345 Scarborough Rd., Briarcliff Manor, NY 10510; balasundar.raju@philips.com), Miachael S. Hughes, Samuel A. Wickline, and Gregory M. Lanza (Washington Univ. School of Medicine, MO 63130)

One of the exciting possible future clinical uses of ultrasound is site-targeted drug delivery using focused acoustic energy. In an earlier study, it was shown that ultrasound exposure promoted lipid delivery from nanoparticles into melanoma cells in culture [Crowder *et al.*, UMB 2005]. In order to extend the study to *in vivo* situations, a novel, combined ultrasound therapeutic and imaging system was developed. The system consisted of a therapeutic transducer (-6-dB BW 1.1 to 2.3 MHz; natural focus 8 cm; f-number 1) with a centrally located and coregistered imaging transducer connected to a commercial ultrasound imaging system. The system provided capability for arbitrary acoustic excitation, electronic axial steering, and mechanical translation of the therapy transducer as well as real-time acquisition of images. The device was tested on bilateral murine tumor models (MDA435). Tumor images provided real-time feedback for treatment planning. The animals were injected with perfluorooctyl bromide nanoparticles targeted to $\alpha_v\beta_3$ integrin and with a radioactive complex. Initial results showed that application of therapeutic ultrasound affected the delivery of the radioactive complex as measured by radioactivity 24 h postinjection. Experimental results and experiences will be

presented detailing the importance of imaging for *in vivo* treatment guidance and planning.

11:45

1aBB5. Ultrasound-assisted penetration of rt-PA into clots to enhance thrombolysis. Saurabh Datta, Sampada Vaidya (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267-0586), Shauna Beiler (VAMC 410, Cincinnati, OH 45220), Gabrielle De Courten-Myer, and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH)

Enhanced rt-PA thrombolysis has been shown in the presence of pulsed ultrasound. Porcine clots were exposed to ultrasound (120 kHz, 80% duty cycle, 1667-Hz PRF, 30 min) and rt-PA at exposure levels exhibiting no cavitation (0.15 MPa), stable cavitation (SC) alone (0.24 MPa), and inertial and stable (IC+SC) combined (0.37 MPa). A separate group of clots was also exposed to rt-PA alone (sham). Percent clot mass loss after each treatment was used to determine thrombolysis efficacy. Cavitation activity was monitored throughout the exposure. After the treatment protocol, penetration of rt-PA and plasminogen was evaluated by immunofluorescene imaging. The highest percent mass loss was observed for SC exposure (26.8%) followed by the IC+SC exposure (19.8%) and the lowest for no cavitation exposure (14.7%). The penetration of rt-PA was distinctly higher for clots exposed to SC compared to unexposed clots. Enhancement of thrombolysis is related to presence of cavitation and SC exposure appears to enhance thrombolysis by facilitating more rt-PA penetration into the clot.

1aBB6. Enhancement of recombinant tissue-plasminogen activator (rt-PA) activity with 2-MHz transcranial Doppler ultrasound. Tyrone M. Porter, Christy K. Holland (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267-0586), Jason M. Meunier, and George J. Shaw (Univ. of Cincinnati, Cincinnati, OH 45267-0769)

A video microscopy technique was developed to monitor the lytic activity of recombinant tissue plasminogen activator (rt-PA) in combination with 2-MHz transcranial Doppler ultrasound (TCD). Human whole blood clots formed on silk sutures were treated with TCD pulses alone [pulse length = 10 cycles, pulse repetition frequency (PRF) = 10.5 kHz, $P_{neg} = 0.18 - 0.47$ MPa], rt-PA alone (3 μg/ml), and rt-PA with TCD pulses in human fresh frozen plasma for 30 min. An inverted optical microscope was used to acquire images of the clot during treatment. The extent of clot lysis was determined by assessing clot width as a function of time. The average percent change in clot width at 30 min was comparable ($\sim$25%) for clots treated with rt-PA and TCD pulses for all output pressures. However, the initial lytic rate (within the first 5 min) was directly related to the TCD pressure output. The initial lytic activity of rt-PA and 2-MHz TCD pulses ($P_{neg} = 0.18$ MPa, PRF = 10.5 kHz, duty cycle = 13%) was similar to the initial lytic activity of rt-PA and 120-kHz ultrasound pulses ($P_{neg} = 0.18$ MPa, PRF = 1667 Hz, duty cycle = 20%).

1aBB7. Ultrasound and biodegradable nanoparticles for anticancer drug and gene delivery. Rinat Esenaliev, Olga Chumakova, Inga Cicenaite (Ctr. for Biomed. Eng., UTMB, Galveston, TX), Shilla Chakrabarty, Mark Evers (UTMB, Galveston, TX), and Valery Andreev (Moscow State Univ., Moscow, Russia)

Poor penetration of macromolecular drugs and genes in tumor cells substantially limits efficacy and safety of cancer chemo- and biotherapy. Cavitation induced by interaction of ultrasound with nanoparticles accumulated in tumors due to the enhanced permeability and retention effect may provide efficient drug and gene delivery in tumor cells. Biodegradable polymer (PLGA) air-filled nanoparticles manufactured in our laboratory were used for enhancement of drug and gene delivery *in vitro* and *in vivo* in human breast, prostate, or ovarian tumors of nude mice (two tumors in each animal). Our specially designed confocal ultrasound system allowed for irradiation and cavitation detection *in vitro* and *in vivo*. It was found that the nanoparticles substantially decrease cavitation threshold and increase cavitation activity *in vitro* and *in vivo*. Moreover, we obtained three-fold increase of tumor cell transfection *in vivo* compared to non-irradiated control tumors of same mice. Substantial tumor necrosis was achieved when macromolecular anticancer drug anti-IGFBP-2 was used in combination with the nanoparticles and ultrasound. Our results indicate that biodegradable polymer nanoparticles and ultrasound can be used for enhancement of drug and gene delivery in tumors. [These studies were supported by the DoD Prostate Cancer Research Program (Grant No. W81XWH-04-1-0247) and Breast Cancer Research Program (Grant No. DAMD17-01-1-0416).]

TUESDAY MORNING, 28 NOVEMBER 2006

IAO NEEDLE/AKAKA FALLS ROOM, 10:30 A.M. TO 12:00 NOON

Session 1aMU

Musical Acoustics: Music Performance

Lydia Ayers, Cochair
Hong Kong Univ. of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Akira Nishimura, Cochair
Tokyo Univ. of Information Sciences, Dept. of Media and Cultural Studies, Faculty of Informatics, 1200-1 Yatoh-cho, Wakaba-ku, Chiba 265-8501, Japan

Contributed Papers

1aMU1. Anticipatory motion in piano performance. Werner Goebl and Caroline Palmer (Dept. of Psych., McGill Univ., 1205 Dr. Penfield Ave., Montreal, QC H3A 1B1, Canada, werner.goebl@mail.mcgill.ca)

Recent motor control studies of piano performance address the anticipatory movements of pianists fingers. The reported study investigated anticipatory motion under manipulations of performance rate or tempo. It was tested whether faster tempi require larger anticipatory preparation, as suggested in piano pedagogy literature. Sixteen skilled pianists repeatedly performed short isochronous melodies from memory at four different tempi (from 500 to 143 ms inter-onset intervals). A passive 3-D motion capture system tracked the movements of 40 markers on the hand, fingers, and the piano keys. The melodies manipulated both the repositioning of certain fingers (thumb and pinkie finger) on the keyboard and the distance between repeating finger movements (three or six tones). Functional data analysis techniques were applied to the analysis of motion trajectories. Data analyses are presented that indicate that the repositioning movement of a finger toward its next goal (keypress) begins sooner at faster tempi than at slower tempi in the x plane (sideways motion on the keyboard).

Applications to piano practice will be discussed. [Work supported by the Austrian Science Fund, NSERC, and the Canada Research Chairs program.]

1aMU2. Effects of altered auditory feedback on piano performance. Noriyuki Takahashi (Kyoto City Univ. of Arts, 13-6 Kutsukake-cho, Oe, Nishikyo-ku, Kyoto, 610-1197, Japan, m03904@kcua.ac.jp) and Minoru Tsuzaki (Kyoto City Univ. of Arts, Kyoto 610-1197, Japan)

Some studies in the field of developmental psychology have suggested that a complex task executed by an expert includes the automated and sequential execution of some subordinate skills. In this study, we regard the development of musical performance as the length of motor control of a musical instrument carried out with a feedforward process. We then verify this assumption through a performance experiment using skilled and unskilled pianists. Both pianists play one musical figure repeatedly in the experiment. In the middle of the performance, a mapping function between the key velocity and its sound output is altered. It is expected that

pianists would try to correct their performance when receiving this altered auditory feedback. By checking the point at which they start to correct, we can infer the length of the unit used with the feedforward process. It is expected that the correction would occur at a specific point in the musical figure if skilled pianists have a certain length unit of automated motor control. On the other hand, the point of correction would have less consistency in the case of unskilled pianists. [Work supported by the Grant-in-Aid for Scientific Research A, No. 16200016, JSPS.]

11:00

1aMU3. An analysis of note deviation manner in piano music. Yukiko Toda, Yosuke Nakayama, Takuya Okura, Naoto Yasuda, and Shigeki Okawa (Chiba Inst. of Technol., 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan)

Human music performance usually contains "out-of-score" features such as players' physical characteristics and emotional expressions besides the on-score information (notes, rests, and various marks). Although those features are not written on scores, they have a very important role to add an artistic sense to the music. As one of the "out-of-score" features, the deviation manner of each note between the score and the human performance is focused on. In this study, among various genres and forms of music, F. Chopin's mazurkas (a form of Polish folk dance), which have triple time and the accent on the second or third beat, are used. By limiting the music genres, it becomes simplified to compare the tendency of players. For the experimental dataset (three pianists, 51 mazurkas), after extracting the fluctuations (time differences) from the regular position of each note, the Fourier transform is applied to the sequence. As a result, a significant difference among the players and the playing modes is obtained, which show the temporal and the rhythm patterns. As a future work, by using the patterns, it is possible to create a virtual performer who has/imitates a certain player or a certain playing mode.

11:15

1aMU4. A technical look at swing rhythm in music. Kenneth A. Lindsay and Peter R. Nordquist (Southern Oregon Univ., 1250 Siskiyou Blvd., Ashland, OR 97520, nordquip@sou.edu)

We investigate swing rhythm in music using computer analysis techniques. Swing is not a genre of music, rather a style of performance. The same musical piece (data) can be played in swing or straight time. Musical notes and structure are identical in both performances, but the notes temporal patterns have slight, significant differences between straight and swing performances. We demonstrate a technical approach analyzing these differences, and show examples of several styles of swing: American Swing, Brasilian Samba, Jamaican Reggae. Prior research into swing rhythm discovered swing ratio representing the temporal variation of the music as performed compared to the music as written. Our work shows that in some cases swing ratio is inadequate to properly model the real patterns of temporal variation. We have extracted details showing swing exists hierarchically, i.e., the patterns of temporal variation played at one time scale may differ from the types of patterns of temporal variation played at a slower or faster time scale. We also show examples of different instruments playing swing differently from each other, locking together at certain canonical time locations that represent standard counting and subdivision of the musical meter. This latter phenomenon has been previously investigated as ensemble swing.

11:30

1aMU5. A jazz arrangement system yielding a guitar chord-form sequence. Koji Sawayama, Norio Emura, Masuzo Yanagida (Doshisha Univ., 1-3, Tatara-Miyakodani, Kyo-tanabe, Kyoto, 610-0321, Japan, dtf0751@mail4.doshisha.ac.jp), and Masanobu Miura (Ryukoku Univ., Otsu, 520-2194, Japan)

Many guitar scores of rock or pop music arranged in jazz style are available all over the world. The amount of software for arranging given sequence of chord names and books of arrangement in jazz style has been growing for these past decades. Using this software, even amateur guitarists can play their favorite tunes in jazz style without knowledge of jazz theory. Most of these software systems, however, do not take musical appropriateness into account. So, sounds of chord-form sequence output by these software systems often mismatch the input melody. Proposed here is a system that generates a chord-form sequence in jazz style for a given set of melody and chord-name sequence. The system analyzes an input set of given melody and chord-name sequence based on jazz theory, and yields the chord-form sequence considering the melody line. It is confirmed that the chord-form sequence yielded by this system meets user requirements and is more preferable than that by software systems on the market. [Work supported by Academic Frontier Project, Doshisha Univ., Knowledge Cluster Project, MEXT, Grant-in Aid for Scientific Research (No.16700154), MEXT, HRC, Ryukoku Univ.]

11:45

1aMU6. A modular system yielding jazz-style arrangement for a given melody and sequence of chord names. Norio Emura, Masuzo Yanagida (1-3, Tatara-Miyakodani, Kyo-Tanabe, Kyoto, 610-0321 Japan, etf1702@mail4.doshisha.ac.jp), and Masanobu Miura (Otsu 520-2194, Japan)

There are many music systems available on the market, such as ones for automatic arrangement of music pieces given as note sequences for solo piano into a piano score in a specific style. Those systems, however, are usually designed to generate music by concatenating existing arrangement patterns, so we cannot expect those systems to meet user requirements. This paper proposes a system that arranges a given set of melody expressed as a note sequence and a chord name sequence into a piano score in modern jazz style based on so-called "jazz theory." The system is implemented in the classical theory of harmony and arrangement techniques described in jazz theory in a modular system. Proposed is a system that gives arrangement by integrating several modules according to structure of input music and user requirements. Performance of the proposed system is evaluated by comparing results given by the proposed system with those given by popular arrangement systems available on the market. Experimental results show that arrangement by the proposed system is significantly superior to arrangement by other systems available on the market. [Work supported by Academic Frontier Project, Doshisha Univ., Knowledge Cluster Project, MEXT, Grant-in Aid for Scientific Research (No. 16700154), MEXT, HRC, Ryukoku Univ.]

Session 1aSC

Speech Communication: Communicative Speech Synthesis and Spoken Dialog

Alan Black, Cochair

Carnegie Mellon Univ., Language Technologies Inst., 5000 Forbes Ave., Pittsburgh, PA 15213

Yoshinori Sagisaka, Cochair

Waseda Univ., GITI, 297 Building, 1-3-10 Nishi Waseda, Shinjuku-ku, Tokyo, 169-0051, Japan

Chair's Introduction—10:30

Invited Papers

10:35

1aSC1. Hidden Markov model-based speech synthesis as a tool for constructing comunicative spoken dialog systems. Keiichi Tokuda (Dept. Computer Sci. and Eng., Nagoya Inst. of Technol., Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan)

For constructing spoken dialog systems which can realize natural human-computer interaction, speech synthesis systems are required to have an ability to generate speech with arbitrary speaker's voice and various speaking styles and/or emotional expressions. Although state-of-the-art speech synthesis systems based on unit selection and concatenation can generate natural-sounding speech, it is still difficult to synthesize various voices flexibly because they need a large-scale speech corpus for each voice. In recent years, a kind of corpus-based speech synthesis system based on hidden Markov models (HMMs) has been developed, which has the following features: (1) original speaker's characteristics can easily be reproduced because all speech features, not only spectral parameters but also fundamental frequencies and durations, are modeled in a unified framework of HMM, and then generated from the trained HMMs themselves; (2) using a very small amount of adaptation speech data, voice characteristics can easily be modified by transforming HMM parameters by a speaker adaptation technique used in speech recognition systems. From these features, the HMM-based speech synthesis approach is expected to be used as a tool for constructing communicative spoken dialog systems: keeping such a viewpoint in mind, basic algorithms and techniques for HMM-based speech synthesis are reviewed.

10:55

1aSC2. Modeling prosody in speech processing. Yang Liu (Computer Sci. Dept., Univ. of Texas at Dallas, MS EC-31, Box 83068, Richardson, TX 75083-0688)

Prosody is an important aspect of speech. In spoken language processing, effective prosody modeling helps to identify additional information beyond words (i.e., "how it is said" instead of "what is said") and thus better understand speech. In this talk, we will discuss how prosodic information is utilized in various speech processing tasks. Prosodic features are extracted to represent duration, pitch, and energy, with different normalization, and modeled using machine learning techniques. Research has shown that prosody provides valuable information for tasks such as automatic identification of important events in spoken language (e.g., sentence boundaries or punctuation, disfluencies, discourse markers, topics, and emotions in dialog). These phenomena are important for enriching speech recognition output and helping downstream language processing modules. Modeling prosodic variation across speakers is also useful for these tasks, as well as for developing speaker recognition systems. Additionally, some issues in machine learning techniques in prosody modeling will be discussed. Understanding how prosody is used to signal interesting events in speech will help to build better synthesis models for generating more natural and expressive speech.

11:15

1aSC3. Expressive speech synthesis using multilevel unit selection. Esther Klabbers and Jan van Santen (Ctr. for Spoken Lang. Understanding, OGI School of Sci. & Eng., OHSU, 20000 NW Walker Rd., Beaverton, OR 97006, klabbers@cslu.ogi.edu)

Generating natural sounding and meaningful prosody is a central challenge in text-to-speech synthesis, especially when generating expressive speech. Recently, we proposed a multilevel unit sequence synthesis approach, based on the general superpositional model of intonation, which describes a pitch contour as the sum of component curves that are each associated with different phonological levels, specifically the phoneme, foot, and phrase. During synthesis, segmental perturbation curves, accent curves, and phrase curves are extracted from the acoustic signal and are combined into target pitch curves; these target curves are then imposed on the acoustic unit sequences using standard pitch modification methods. This approach represents an attempt to combine the strengths of the two dominant approaches to speech synthesis: unit selection synthesis, which preserves all details of natural speech but struggles with coverage of the very large combinatorial space of phoneme sequences and prosodic contexts, and diphone synthesis, which addresses coverage by generating rule-based synthetic target prosody and imposing it on acoustic units using signal modification methods. This approach minimizes prosodic modification artifacts, optimizes the naturalness of the target pitch contour by using quasi-natural contours, yet avoids the combinatorial explosion of standard unit selection synthesis.

1aSC4. Realization of rhythmic dialogue on spoken dialogue system using paralinguistic information. Shinya Fujie and Tetsunori Kobayashi (Dept. Comput. Sci., Waseda Univ., 3-4-1 Okubo, Rm. 55N-509, Shinjuku-ku, Tokyo 169-8555, Japan)

Paralinguistic information recognition and its application to spoken dialogue system are discussed. Paralinguistic information brings various useful cues that cannot be obtained from linguistic information. The system may know, for example, who should speak next, how the user really feels, and so on. In this paper, we implement several paralingusitic information recognition systems, such as prosody-based attitude recognition system, back-channel feedback generation/recognition system, head gesture recognition system, and so on. Finally, we combine them into the spoken dialogue system on the humanoid robot, and realize a rhythmic dialogue.

TUESDAY MORNING, 28 NOVEMBER 2006

WAIALUA ROOM, 10:30 A.M. TO 12:00 NOON

Session 1aUW

Underwater Acoustics: Time Reversal Methods

Mark Stevenson, Cochair
NATO Saclantcen Research Centre, CMR 426 Box 714, APO, AE 09613-0714

Shunji Ozaki, Cochair
Oki Electric Industry Co., Ltd., 7-12 Toranomon, 1-Chome, Minato-Ku, Tokyo 105-8460, Japan

Contributed Papers

10:30

1aUW1. The focused acoustic fields time reversal experiment series. Mark Stevenson, Piero Guerrini (NATO Undersea Res. Ctr., La Spezia, Italy), Tuncay Akal, Heechun Song, Karim Sabra, Philippe Roux, William S. Hodgkiss, and William A. Kuperman (Scripps Inst. of Oceanogr., La Jolla, CA)

We review a progression of at-sea experiments that brought the concept of underwater, acoustic time-reversal (TR) focusing in the sea from an unverified hypothesis to reality. When the experiments began, it was completely unknown whether the time-varying nature of the oceanic environment would support sufficient temporal stability for the method to succeed, even at long ($\lambda > 3$ m) acoustic wavelengths. A 10-year progression of experiments showed that not only was this possible, but that various applications could be implemented using higher frequency sound (up to 18 kHz), including phase coherent digital communications. Our most recent experiment, FAF-06, investigated multiuser applications for passive TR communications using small transmit and receive arrays. We demonstrated throughput of 10 kbps at several km range in a shallow-water propagation channel where the transfer function had an ISI span of several hundred symbols. [Work supported by NATO and ONR.]

10:45

1aUW2. Beam sweeping in sediment layer using phase-conjugated pseudo sound sources. Yoshiaki Tsurugaya, Toshiaki Kikuchi[a], Ryoichi Iwase[b], and Koich Mizutani[c] (NEC Corp., 1-10 Nishinchou Fuchu-city, Tokyo, 183-8501, y-tsurugaya@bp.jp.nec.com)

When an object that is buried in a sediment layer is detected from the diagonal, detection is hampered by reflection from the bottom surface. Stated in different terms, the sound-wave energy that reaches the object is small because of the reflection and refraction that occurs at the bottom surface when a sound wave is applied to the object in the sediment for diagonal detection. To alleviate that problem, a pseudo sound source can be set up in sediment layer. The sound wave radiated from the pseudo sound source is then received by a time-reversal array. The sound wave can be injected into the sediment layer by reversing the received signal with time and radiating it again from the time-reversal array. The beam can hit the target through movement of this pseudo sound source along the bottom surface. The sound wave that hits the target is reflected and returns in water. The sound wave is subsequently reflected many times at the

surface and bottom and is diffused over time. The passive-phase conjugate processing is then given to the received diffusing signal to reduce this diffusion. The pulse is compressed by this processing, thereby obtaining the target position. [a]Currently at NDA; [b]Currently at JAMSTEC; [c]Currently at Univ. of Tsukuba.

11:00

1aUW3. Synchronized time-reversal focusing with application to virtual array formation, beam forming, and remote imaging. Shane C. Walker, Philippe Roux, and W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093)

By sampling and time reversing the field from a distant probe source, a time-reversal mirror (TRM) can produce an intense acoustic focus at the probe source location. Extending this idea to a collection of sources, this work demonstrates that a TRM can generate a virtual array at a downrange location, which can act as a remote platform for time delay beam forming and other coherent wavefront techniques useful for imaging in the vicinity of the virtual array. The technique requires no *a priori* information about the environment, only the knowledge of the source fields and a means for synchronizing them. In one promising application the TRM is used to coherently image an extended region around a distant passive array. The geometry of the source array is arbitrary. Recent advances in self-adaptive acoustic Green's function determination may provide opportunities to apply this method without need for a source array. Simulation and experimental results are discussed.

11:15

1aUW4. Robust wideband time-reversal using time-domain principal component analysis. W. J. Higley, Philippe Roux, W. S. Hodgkiss, W. A. Kuperman, Heechun Song, Tungay Akal (Marine Physical Lab., SIO, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0238), and J. Mark Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

Time reversal (TR), a two-step process, involves time reversing the measured transfer functions from a source and then retransmitting this signal to produce a focus back at the source position. When attempting TR in a fluctuating ocean environment, it is desirable to alter the transmitted signal such that it produces a more robust focus. This has typically been done with work based on frequency domain analysis. In contrast, this

work, using oceanic data measured during FAF05, is broadband and the development is totally in the time domain. Different definitions of robustness give different optimal answers to the question of how to design the most robust transformation of the time-reversed data of the first step in the TR process. Two definitions of robustness are studied, one related to outage probability and another related to average power. For each definition, a different approach to a practical approximation is taken. In the case of outage probability, we iteratively solve a minimax optimization problem. In the case of average power, principal component analysis is performed on time-domain data.

11:30

1aUW5. Identification and classification of resonant targets with iterative, single-channel time reversal in a noisy environment. Zachary J. Waters, R. Glynn Holt, Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215), and Benjamin R. Dzikowicz (NSWC-Panama City, Panama City, FL 32407)

The identification and classification of mines resting on and buried beneath the sea floor is an important and challenging problem. Previous numerical results indicate that iterative time reversal using a single-element transducer increases the signal-to-noise ratio of the backscattered return while simultaneously converging to a narrow-band signal characteristic of the scattering target [D. M. Pierson, "Buried object detection using time-reversed acoustics," Ph.D. dissertation, North Carolina State University 2003]. This technique offers an inexpensive and simple means for maximizing signal level in a noisy or cluttered environment. We investigate this phenomena through experiments performed in a water tank using a 6.35-mm-diam hollow steel sphere and broadband $Q \sim 1.5$) piston transducers operating in the 100-kHz to 2-MHz frequency range. Time-reversal convergence of the interrogation pulse was readily achieved using a variety of broadband (0.1–2 MHz) pulses to initiate the process. Results

indicate that the spectrum of the echo rapidly converges to the dominant resonant mode of the target, with a significant enhancement in signal level even in the presence of broadband noise and coherent clutter returns. [Work supported by Office of Naval Research Award No. N000140610044.]

11:45

1aUW6. Experiment of time-reversal communication in deep ocean. Takuya Shimura, Yoshitaka Watanabe, and Hiroshi Ochi (JAMSTEC/MARITEC, 2-15, Natsushima-cho, Yokosuka-City, 237-0061, Japan, shimurat@jamstec.go.jp)

The basic study on the long horizontal acoustic communication using time-reversal waves have been carried out by the authors, and the method of combining time reversal and adaptive filter was proposed to enhance the communication accuracy and to achieve the communication with a moving source-receiver. This time-reversal technique will be applied to the communication in the deep ocean with an autonomous underwater vehicle, "Urashima," which is being developed in JAMSTEC. The experiment was executed to demonstrate the effectiveness of the proposed method in the area at the depth of 1000 m in Suruga-bay. In the experiment, the 500-Hz projectors were used, and the range between the focus and the time-reversal array (TRA) was 10 km. The measurement system of the TRA side was moored with the underwater winch to compose the virtual array by changing its depth. Unfortunately, the projector of the TRA side had a breakdown, so it was abandoned to execute active time-reversal communication. However, experiments of passive time-reversal communication could be achieved. It was verified that the proposed method makes it possible to assure the communication channel with the TRA that consists of 20 elements. [This work was supported by MEXT KAKENHI 16686050.]

TUESDAY AFTERNOON, 28 NOVEMBER 2006   LANAI ROOM, 1:00 TO 6:15 P.M.

Session 1pAA

Architectural Acoustics: Computer Modeling of Room Acoustics II

Lily M. Wang, Cochair
Univ. of Nebraska–Lincoln, 200B Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681

Shinichi Sakamoto, Cochair
Univ. of Tokyo, Inst. of Industrial Science, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. Numerical analysis of sound propagation in rooms using the finite difference time domain method. Shinichi Sakamoto, Ayumi Ushiyama, and Hiroshi Nagatomo (Inst. of Industrial Sci., The Univ. of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan, sakamo@iis.u-tokyo.ac.jp)

Impulse response is an important quantity for room acoustics. To predict the impulse response, several types of calculation methods based on wave theory, such as the boundary element method (BEM), finite element method (FEM) and finite difference method (FDM), are being investigated. Among them, the authors are exploring application of the finite difference time domain (FDTD) method to calculate room sound fields. Using the method, by which the acoustical quantities at discrete grid points of a closed field are successively solved step-by-step according to vector operations, the impulse responses can be obtained directly in a time domain using little computer memory storage. Furthermore, changes of spatial distributions of room sound fields in time can be obtained easily by storing the sound pressure at grid points and displaying them by animation. In this presentation, several examples of the FDTD analyses of room impulse responses are reported and computer animation of their room sound fields is introduced. In addition, examination of the applicability of prediction of the impulse responses by comparison with *in situ* measurement of a small hall is discussed.

1:25

1pAA2. Applications of large-scale finite-element sound-field analysis utilizing Architectural Institute of Japan-benchmark platform. Toru Otsuru, Noriko Okamoto, and Takeshi Okuzono (Faculty of Eng., Oita Univ., 700 Dannoharu, Oita 870-1192, Japan, otsuru@cc.oita-u.ac.jp)

A computational system, namely large-scale finite-element sound-field analysis(LsFE-SFA), has been developed by the authors. It is based on the finite-element method (FEM) with an iterative solver of linear equation that can reduce memory requirement of the system drastically. To show the effectiveness of LsFE-SFA, sound fields in a reverberation room and in a music hall were analyzed. The problem settings are provided by the web site of ''AIJ-Benchmark Platform for Computational Acoustics,'' which is run by a subcommittee of Architectural Institute of Japan (AIJ). First, sound-pressure level distribution maps at about 300 points in the reverberation room computed by LsFE-SFA were compared with those by measurement to show the accuracy of LsFE-SFA to be within 1-dB difference. The frequencies investigated were in 1/3-octave bands with the center frequency raging from 63 to 250 Hz. In the computations, boundary absorption is modeled by use of normal impedance that was measured *in situ* by use of the authors' newly proposed method, or EA-Noise method. Next, the diffuser shapes on the walls in the music hall were examined by LsFE-SFA to give better diffuseness. The investigations showed the effectiveness of LsFE-SFA.

1:45

1pAA3. The application of boundary integral equation methods to the numerical calculation of transient waves in rooms. Yasuhito Kawai (Dept. of Architecture, Kansai Univ., 3-35, Yamatecho 3, Suita, Osaka, 564-8680 Japan)

The boundary integral equations derived from Kirchhoff's formula and its normal derivative form are available for the numerical calculation of transient waves in rooms. The two equations are compared in view of the stability and accuracy in the numerical calculation of responses in a rigid cubic room in which rigorous responses can be obtained by the image method. Some source waves and some problems appear in the numerical implementation such as the singular integrals, and the treatment of absorbent surfaces is also investigated. Finally, it is shown that the boundary integral equation method is a powerful tool to obtain transient waves in rooms.

2:05

1pAA4. Fogg Art Museum lecture room: A calibrated recreation of the birthplace of room acoustics. Brian F. G. Katz (Percept. Située, LIMSI-CNRS, BP 133, F91403 Orsay, France, brian.katz@limsi.fr) and Ewart A. Wetherill (28 Cove Rd., Alameda, CA 94502))

At the end of the 19th century, Wallace Clement Sabine undertook the task of correcting the acoustics of the Fogg Art Museum lecture hall. His work on this acoustically difficult (semi-circular, domed, arched) auditorium was the basis for his monumental work on reverberation time and absorption, beginning the science of room acoustics as it is now known. The room underwent various renovations over 70 years, concluding with its demolition in 1973. Just prior to this event, one author was fortunate to have measured several room impulse responses. Through the use of architectural plans and the few existing photos, a computational room acoustical model was created. This model was calibrated using the historically important measurements as a reference. The geometrical model incorporates the room's various architectural phases. Using computer auralizations, the room is now accessible in acoustical terms, with the ability to ''perience'' an educated rendering of the acoustics of the room as Sabine found it in 1895. The results of this study, particularly the geometrical model, are intended for public use and are made generally available for room acoustics students and researchers who chose to follow in Sabine's footsteps (www.catt.se/FoggArtMusum.htm). Portions of this presentation were presented at Forum Acousticum 2005.

2:25

1pAA5. Auralization of an orchestra using multichannel and multisource technique. Michelle C. Vigeant, Lily M. Wang (Architectural Eng. Prog., Univ. of Nebraska—Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, mvigeant@mail.unomaha.edu), and Jens Holger Rindel (Tech. Univ. of Denmark, DK-2800 Kgs. Lyngby, Denmark)

Previous work has shown the importance of including source directivity in computer modeling for auralizations. A newer method to capture source directivity in auralizations is the multichannel technique, which uses multichannel anechoic recordings. In this study, five-channel anechoic recordings were obtained for every orchestral part of two symphonies at the Technical University of Denmark. Five-channel auralizations were then created for each instrument, located at its typical position on-stage in a concert hall, by convolving five impulse responses from sources that each represent approximately one-fifth of a sphere with the corresponding anechoic recording channel and mixing these together. The multichannel auralizations from each instrument were subsequently combined to produce a full multichannel auralization of the entire orchestra. The results from listening tests will be presented, illustrating the perceived changes in realism, source width and depth, and source separation between the output of this full orchestral mix of multichannel auralizations and other auralization methods, including the following: (*i*) a mix of single-channel auralizations from each instrument, and auralizations made with a single channel orchestral anechoic recording using (*ii*) a surface source and (*iii*) a single omni-directional source. [Work supported by the National Science Foundation.]

2:45

1pAA6. Prediction of scattering coefficients for use in room-acoustic simulation. Wolfgang Ahnert, Stefan Feistel, and Steffen Bock (Ahnert Feistel Media Group, Berlin, Germany)

While absorption coefficients have become easily available and published worldwide, computer modeling of room acoustics suffers until today under the lack of data for scattering surfaces. As a result, in most cases rules of thumb have to be used. In this work a new approach of predicting scattering coefficients is introduced. Based on the geometrical parameters of the surface, the complex reflection properties are calculated utilizing an elementary-wave method. In this respect, computed data are compared afterwards with measurement data and its applicability for use in acoustic simulation is investigated.

3:00

1pAA7. A comparison of computer modeling auralization and binaural room recordings for a multipurpose high school auditorium utilizing an orchestral shell, overhead panels, and clouds.. Ron Freiheit (Wenger Corp., 555 Park Dr., Owatonna, MN 55060)

The development of auralization techniques has provided a new method to demonstrate potential acoustical benefits of changes in design or the application of acoustic treatments in specific spaces. It's useful to compare the results from the auralization process with actual binaural recordings in the physical space to discover similarities and differences. A typical high school proscenium-style, multipurpose auditorium was modeled, including an orchestral shell, overhead panels, and clouds. Auralizations were made of the auditorium at various locations in the audience area and on stage, with and without the shell and overhead panels, in place. These were compared with binaural recordings made at the same locations within the auditorium. Auralization metrics will be compared with the binaural recordings, highlighting the changes made with the shell and overhead panels in place, along with the similarities and differences between the two methods.

3:15–3:30   Break

3:30

1pAA8. Optimal dispersion/dissipation-error finite-difference method for room impulse response simulation. Yusuke Naka (Dept. of Aerosp. and Mech. Eng., Boston Univ., 677 Beacon St., Boston, MA 02215, ynaka@bu.edu), Assad A. Oberai (Rensselaer Polytechnic Inst., Troy, NY 12180), and Barbara G. Shinn-Cunningham (Boston Univ., Boston, MA 02215)

The finite difference method is suitable for the calculation of acoustic impulse responses because of its computational simplicity and efficiency. However, for room acoustics applications, where reverberation makes it necessary to calculate the room impulse response and acoustic energy over a long duration, most standard finite-difference formulations are not practical because of the accumulation of dispersion and dissipation errors. In order to calculate very long impulse responses, a new, efficient, high-order finite-difference method is developed in this study. The new method is optimized to minimize the total dispersion and dissipation errors and hence the error accumulated in each time step resulting from discretization in both space and time. The error of the new method is significantly reduced so that the impulse response can be calculated over the whole reverberation time (about 1 to 3 s), using a relatively coarse spatial (6 grid points per wavelength) and a moderate temporal (CFL = 1.0) resolutions. [Work supported by AFOSR.]

3:45

1pAA9. Temporal decay of broadband sound fields in enclosures using a boundary element method. Donald Bliss, Ben Manning (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, donald.bliss@duke.edu), and Jerry Rouse (Sandia Natl. Labs, Albuquerque, NM 87185-0847)

A method to predict time-dependent behavior of sound fields in enclosures with diffuse reflection boundaries is described. A boundary element formulation utilizes broadband energy-intensity boundary sources, including propagation time delays. Temporal behavior is expressed in terms of a higher-dimensional eigenvalue/eigenmode problem, with source strength distributions expressed as eigenmodes. A temporal interpolation scheme is used to reexpress the actual delays as a set of integer-multiple delays. The interpolation scheme produces a discrete set of primary eigenvalues and eigenvectors, the total number of which is an integer multiple M of the number of panels N. The value of M is set by the number of interpolation subdivisions of the enclosure wave transit time. For decay from steady-state conditions, the initial value problem involves specifying the initial panel strengths, and $M-1$ time derivatives of each panel strength set to zero. In temporal decay from steady state, solutions exhibit rapid short-time spatial redistribution of energy, followed by long-time decay of a predominant spatial eigenmode. Long-time decay depends on the behavior of the most slowly decaying eigenmode and the relative source-panel strengths do not depend on initial conditions. Short-time adjustment and decay depends on initial source characteristics and the distribution of wall absorption.

4:00

1pAA10. Boundary element prediction of steady-state broadband interior sound fields using absorption-based scaling. Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, donald.bliss@duke.edu) and Jerry Rouse (Sandia Natl. Labs, Albuquerque, NM 87185-0847)

Energy-intensity boundary elements are used in combination with the method of absorption-based scaling to predict steady-state broadband sound fields in enclosures with diffuse reflection boundaries. The wall absorption is expressed in terms of an overall absorption parameter—the spatially averaged value—and spatial variations around this mean. Boundary element strengths are expressed in a power series of the overall absorption, treated as a small parameter, thereby giving a separate problem at each order. The first problem has a uniform mean-square pressure proportional to the reciprocal of the absorption parameter. The second problem gives mean-square pressure and intensity distributions that are independent of the absorption parameter and are primarily responsible for the spatial variation of the reverberant field. This problem depends on the location of sources and the spatial distribution of absorption, but not overall absorption. Higher order problems proceed in powers of the absorption parameter, as corrections to the primary spatial variation. Boundary element solutions are obtained for each order, and the simplest model for spatial variation gives fairly accurate predictions. The method gives physical understanding of the causes of the spatial variation of enclosure sound fields and provides specific insights into the behavior and design of acoustic spaces.

4:15

1pAA11. Comparison of sound fields in rooms computed by finite-element method and by cone beam method. Noriko Okamoto, Toru Otsuru, Reiji Tomiku (Oita Univ., 700 Dannoharu, Oita-city, Oita 870-1192, Japan, nokamoto@cc.oita-u.ac.jp), Takeshi Okuzono, Kouta Isobe (Oita Univ., Oita, Japan), and Hiroshi Furuya (Kyushu Kyoritsu Univ., Fukuoka 807-8585, Japan)

Sound fields in realistic rooms were simulated by different kinds of computational methods, and comparisons were conducted. One is based on the wave acoustics and the other is based on the geometrical acoustics, i.e., the finite-element method (FEM) and the conical beam method

(CBM). As for the software applied here, the authors have developed their own FEM code, namely LsFE-SFA, to conduct FEM computations, and they employed commercial software, i.e. RAYNOISE, to conduct CBM computations. First, theoretical basis of LsFE-SFA was given briefly. Next, sound-pressure distributions in a reverberation room with the volume of 165 cubic meters were computed by both methods to compare the agreement with measured sound-pressure levels. The boundary conditions due to the sound absorption were changed to examine the relation between the absorbent condition and the agreements. The agreement between FEM and measurement were found to be better throughout the examination. Next, both FEM and CBM were applied onto the analysis of large-scale sound fields in a multipurpose hall with the volume of 37 000 cubic meters to discuss the agreement and difference in detail.

4:30

1pAA12. Absorption and aperture size effects on subjective response to complex computer modeled coupled volume systems. David T. Bradley (BRC Acoust. & Technol. Consulting, 1741 1st Ave. S, Ste. 401, Seattle, WA 98134, dbradley@brcacoustics.com) and Lily M. Wang (Univ. of Nebraska-Lincoln, Omaha, NE 68182-0681)

A typical coupled volume concert hall consists of a large, absorbent main volume linked to a smaller, reverberant auxiliary space through a series of acoustically transparent openings. This type of space can exhibit a nonexponential sound decay, a phenomenon that is referred to as the double slope effect (DSE). An imagined, realistic coupled volume concert hall has been computer modeled, and room impulse responses of the space have been obtained. Several configurations of the hall were conceived in which the absorption and aperture size were modified. Paired-comparison tests were conducted to determine listener preference for the sound fields developed from auralizations of the configurations. The results from this subjective testing were statistically analyzed using multidimensional scaling methods. The analysis showed that listener preference increased with decreasing DSE, with highest preference levels for low and medium DSE. These results will be presented and contrasted with objective parameters calculated from the room impulse responses of the computer-modeled configurations.

4:45

1pAA13. Numerical analysis for attenuation due to sound absorption in small and large closed spaces using finite-element method. Takao Yamaguchi (Dept. of Mech. Systems Eng., Gunma Univ., 1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan), Junichi Kanazawa (Musashino Damping Technol., Tachikawa 190-0002, Japan), and Katsuji Akamatsu (Machinery Acoust., Nada-ku, Kobe 657-0015, Japan)

A finite element for media having sound absorption is derived and a code is developed. For this element, an unknown parameter is pressure, and both effective density and bulk modulus of the media have complex quantity to represent damped sound fields. This model for the media having sound absorption was adopted by H. Utsuno's research group for their boundary element formulation [Utsuno et al., AIAA J. **28**(11), 1870–1875 (1990)]. Firstly, a three-dimensional sound field in a small closed space ($306 \times 255 \times 204$ mm^3), containing a sound-absorbing material, is calculated using the finite element method. Validity of the calculated response is verified by comparison with the corresponding experimental result carried by H. Utsuno. Secondly, a two-dimensional sound field in relatively large closed space (about 7×3 m^2) is calculated varying damping of its medium. The shape of the large space is rectangle, and the rectangular space has a roof. Distribution of sound pressure is calculated when the velocity of the roof is given as white noise. Under the low damping condition for the medium, distribution of sound pressure is almost uniform like the diffused sound field. Under the high damping condition for the medium, sound pressure become larger when the observation point is closer to the roof.

5:00

1pAA14. Real-time audio rendering system for virtual reality. Tobias Lentz, Dirk Schroeder, Gottfried K. Behler, and Michael Vorlaender (Inst. fuer Technische Akustik, RWTH Aachen Univ., D-52056 Aachen, Germany)

Systems of virtual acoustics are more and more coming into the focus of research and application. Creating a virtual sound scene with spatially distributed sources requires a technique for adding spatial cues to audio signals and an appropriate reproduction. A real-time audio rendering system is introduced that combines room acoustics, dynamic crosstalk cancellation, and multitrack binaural synthesis for virtual acoustical imaging. The room acoustics component takes into account the full up-to-date algorithmic approach of specular reflections and scattering, but with real-time processing. The real-time performance of the software was reached by introduction of a flexible framework for the interactive auralisation of virtual environments. The concept of scene graphs for the efficient and flexible linkage of autonomous-operating subscenes by means of so-called portals has been incorporated into the existing framework, combined with an underlying BSP-tree structure for processing geometry issues very fast. Using this framework enables a significant reduction of computation time for both applied algorithms (deterministic image sources and a stochastic ray tracer). This enables the simulation of indoor or outdoor spatial distributed sources and, in addition to that, near-to-head sources for a freely moving listener in room-mounted virtual environments without the use of headphones.

5:15

1pAA15. Effects of peg's positions of low-registered strings on acoustic gain by a stage riser: Theoretical model analysis. Shinsuke Nakanishi (Dept. of Integrated Architecture, Hiroshima Intl. Univ., Hirokoshingai, Kure, Hiroshima 737-0112 Japan, s-nakani@it.hirokoku-u.ac.jp)

This paper theoretically analyzes acoustic gain by a stage riser, which has been observed when playing the low strings, e.g., a violoncello and a double bass, on the riser in concert auditoria [Beranek et al., J. Acoust. Soc. Am. **36**, 1247–1262 (1964)]. The previous paper reported that "the tone quality increased in richness with the risers, particularly when the pegs of the instruments were placed about 4 in. from the edge" as musician's comments. The author discusses effects of the peg's positions of the low strings on the acoustic gain, through numerical examples of theoretical model analysis. The solution of a sound field is obtained when playing the low string on the riser placed on the rigid floor, which receives a spherical incident wave and mechanical excitation force from the low string. Numerical examples show that peg's positions of the low string affect peaks and dips of the acoustic gain, which are caused by eigenvibrations of a top plate of the riser. And, the acoustic gain becomes a positive value at more frequencies as the standing position of the low string shifts towards the riser's edge.

5:30

1pAA16. The program battle at Ganz Hall. Candice Tillery and Jay Pleckham (Columbia College Chicago, 33 E. Congress, Chicago, IL 60605, listen_42@sbcglobal.net)

Current modeling programs offer a wide range of analysis and prediction tools, as well as detailed modeling capabilities and user friendly formats. However, these tools are only useful if they can do their job correctly. The accuracy of the program, its cost, and its relevance to the environment in question are all vitally important when selecting software for a project. This paper addresses these issues through a comparison of two of today's most popular modeling programs: CARA and EASE. An acoustical analysis was performed on Roosevelt University's Ganz Hall, which seats 195 people and is used primarily for student recitals. Data were acquired concerning the hall's reverberation and frequency response. Numerous calculations were performed to determine the room's musical clarity, speech clarity, and early decay time. The hall was then modeled in the two programs in question, and the results were compared to determine each program's accuracy.

5:45

1pAA17. Parallel computing on the sound fields by bread slicing method. Yasuhiro Tokita (Waseda Environ. Res. Inst., Waseda Univ., 1011 Okuboyama, Nishitomida, Honjou, Saitama, 367-0035, Japan), Yasuhiro Oikawa, and Yoshio Yamasaki (Waseda Univ., Honjou, Saitama 367-0035, Japan)

Numerical analyses by the explicit scheme of the finite difference method greatly reduce the number of known variables that are needed to solve an unknown variable in comparison with the implicit methods. This fact means that the amount of communications decreases among each node of the computer cluster on parallel computing. Generally, on the parallel computing of the distributed memory system, the communicating time occupies most of the total calculating time. Therefore, the reduction to the amount of the communication directly causes the reduction of the calculating time. The parallel computing by the explicit scheme of the finite difference method can enable the building up of high-performance algorithms. In order to reduce the amount of communication among each node, one has proposed a method to divide the field of numerical analyses. It is called the ''bread slicing method.'' With this method, the fields must be divided into pieces parallel to the section that has the least area. The results showed that the method in this study enabled the total amount of communication in the computer cluster to be reduced and the geometries of communication to be made simpler.

6:00

1pAA18. Theoretical analysis of statistical properties of the sound-pressure field and sound energy density field in rectangular enclosures. Buye Xu and Scott D. Sommerfeldt (Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

The sound-pressure field and sound energy density (ED) field for rectangular enclosures are obtained theoretically by solving the Helmholtz equation and Euler's equation with low-energy absorption boundaries. Based on those results, both analytic and numerical methods are applied to compute some statistical properties. The results of the statistical analysis show that the spatial variation of the ED field is generally much less than that of the pressure field in reverberant enclosures. In addition, a modified ED quantity is developed and will be discussed.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

KOHALA/KONA ROOM, 1:30 TO 5:30 P.M.

Session 1pAB

Animal Bioacoustics and Underwater Acoustics: Remote Monitoring of Biological Sounds II

Robert Gisiner, Chair
Office of Naval Research, 875 N. Randolph St., Arlington, VA 22203

Contributed Papers

1:30

1pAB1. Humpback whale social sounds: Sources levels and response to playback. Michael J. Noad, Rebecca Dunlop (School of Veterinary Sci., Univ., of Queensland, St. Lucia, Qld, 4072, Australia, mnoad@uq.edu.au), Douglas H. Cato (Defence Sci. and Tech Org., Pyrmont, NSW 2009, Australia), Dale Stokes (Scripps Inst. of Oceanogr., La Jolla, CA 92037), Patrick Miller, and Nicoletta Biassoni (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Humpback whale ''social sounds'' appear to be used in communication when whales interact but they have received little study in comparison to the song. During experiments as part of the Humpback whale Acoustics Research Collaboration (HARC), whales migrating past the study site on the east coast of Australia produced a wide range of social sounds. Whales were tracked visually using a theodolite and singers were tracked acoustically using an array of five widely spaced hydrophones. Source levels of social sounds were calculated from the received level of the sounds, corrected for measured propagation loss. Playbacks of social sounds were made using a J11 transducer and the consequent reactions were recorded in terms of the change in direction of the migrating whales in relation to the playback position. In one playback, a DTAG was place on a female with calf. Playback of social sounds resulted in significant changes in the course of the migrating whales, in some cases towards the transducer while in others it was away from the transducer. From the estimates of source levels it is possible to assess the effectiveness of the playback and the range of influence of social sounds. [Work supported by ONR and DSTO.]

1:45

1pAB2. Megaclicks: Tagged humpback whales produce click trains and buzzes during foraging. Alison K. Stimpert, Whitlow W. L. Au (MMRP/Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734, stimpert@hawaii.edu), Mark P. Johnson (Woods Hole Oceanog. Inst., Woods Hole, MA 02543), David N. Wiley (Stellwagen Bank Natl. Marine Sanctuary, Scituate, MA 02066, USA), and Roland Arsenault (Univ. of New Hampshire, Durham, NH 03824)

Mysticetes and odontocetes differ in many morphological and behavioral aspects related to feeding. Odontocetes echolocate using clicks to localize prey, while the mechanism through which mysticetes orient and find food is unknown. In this study, DTAGs were deployed on four humpback whales (*Megaptera novaeangliae*) observed feeding off the northeastern United States. From two of the four tagged whales, bouts of clicks produced with a temporal pattern resembling odontocete echolocation were recorded and 102 bouts of megaclicks were identified. Spectrally, megaclicks were unlike previously reported humpback sounds or odontocete clicks, and also had much lower amplitudes (peak frequency $\sim$1700 Hz, RL 148 ± 6 dB Re: 1 uPa peak). Interclick intervals (ICI) were bimodally distributed (peaks at 25 and 80 ms), with shorter ICIs occurring at the end of megaclick bouts, in a pattern similar to an odontocete terminal buzz used in foraging. Behavioral sensor data supported a possible foraging function for megaclicks. All megaclick bouts occurred during nighttime hours, and the terminal buzz was associated with a sharp body roll of greater than 70 deg. Low source levels, among other characteristics, contraindicate an echolocation function; however, further analysis and research is required to distinguish among this and other hypotheses for sound function.

3012   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3012**

2:00

1pAB3. Temporal, geographic, and density variations in the acoustic activity of snapping shrimp. Marc O. Lammers, Sara Stieb, Whitlow W.L. Au, T. Aran Mooney (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Russell E. Brainard, and Kevin Wong (Pacific Islands Fisheries Sci. Ctr., Honolulu, HI 96814)

The sounds produced by snapping shrimp (genus *Synalpheus* and *Alpheus*) are the main contributors to the ambient sound field of shallow-water tropical and subtropical marine habitats worldwide. Despite this, relatively little is known about the temporal and spatial variability of snapping shrimp activity. To investigate this issue further, ecological acoustic recorders (EARs) were utilized to obtain time-series records of the ambient sound field from multiple locations in Hawaii and American Samoa. Strong diel trends marked by sharp increases in acoustic activity during crepuscular periods were observed at all sites. However, sites varied greatly both in absolute sound levels and relative fluxes between day and night. In addition, fluctuations in snapping shrimp sound levels were found to track lunar cycles, with higher levels occurring during the new moon. To further investigate the reason for the higher levels observed during crepuscular periods, individual shrimp were localized using a three-hydrophone line array on both a coral reef and a mud and algae habitat in Kaneohe Bay, Hawaii. The density of acoustically active shrimp was found to be higher during the sunset period than during the day or at night.

2:15

1pAB4. An automated approach to detecting pulsive gray whale (*Eschrichtius robustus*) vocalizations within acoustic data to estimate call rate changes over time in Laguna San Ignacio. Melania Guerra, Aaron M. Thode (Marine Physical Lab, SIO UCSD, 9500 Gilman Dr., La Jolla, CA 92037-0238), Marilyn E. Dahlheim (Natl. Marine Mammal Lab, Seattle, WA 98115), Jorge Urban (UABCS, La Paz, Mexico), and Amanda M. Koltz (Cornell Univ., Ithaca, NY 14853)

Acoustic recordings of type 1b gray whale calls were made in San Ignacio Lagoon, during the wintering months of 2005 and 2006 when these animals were present to breed and raise calves. A collection of horizontal, bottom-mounted arrays containing two or three elements monitored the mouth of the lagoon over consecutive days, targeting vocalizations below 1 kHz. Because these pulsive calls are highly variable in the timing and number of pulses, standardized detection methods involving matched-filtering or spectrogram correlation have been unsuccessful. A sequence of automated procedures programed in MATLAB and ISHMAEL was implemented to assist with the detection of calls within the dataset. Simple manual recognition based on human judgment was also performed on the shorter time series to serve as an independent check. Results for hourly call rates were computed for both seasons individually. A similar such acoustic dataset had been previously collected by Dahlheim (1987) in the years 1982–84 and analyzed manually. This dataset would be digitized and examined with the new automated analyses in order to compare changes in average call rates over the two-decade interval. A rough relationship with animals present can be drawn, as census numbers are available for both periods.

2:30

1pAB5. Characteristics of the deep scattering layer in the Gulf of Mexico as they relate to sperm whale diving and foraging behavior. Alyson J. Azzara (Dept. of Oceanogr., Texas A&M Univ., College Station, TX 77843-3146, alygirl@tamu.edu), Aaron Thode (Scripps Inst. of Oceanogr., 9500 Gilman Dr., San Diego, CA 92109), Tom Norris (Sci. Applications Intl. Corp., San Diego, CA), Kelly Benoit-Bird, and Douglas C. Biggs (Oregon State Univ., Corvallis, OR 97331)

In June 2005 we used a 38-kHz ADCP and two Simrad fishery echo-sounders to collect backscatter data and passive acoustic array recordings of sperm whale sounds, as part of the Sperm Whale Seismic Study (SWSS). Dive profiles and presumed foraging depths of the whales were derived using a 3-D tracking technique [see abstract by Thode and Norris (this conference)]. Relationships are examined between dive profiles and acoustic backscatter intensity measurements from the main deep scattering layer (DSL) at a daytime depth of 400–600 m. Passive detections of creaks, believed to be associated with prey acquisition, are merged with dive profiles to provide insight between mid-water backscatter returns and foraging effort. Results to date indicate whales diving in three depth ranges: deeper than 800 m, between 500 and 800 m, and less than 500 m. Dive patterns indicate some whales dive through the main DSL into deeper secondary layers seen by the 38-kHz ADCP. There seems to be no difference in dive patterns between night and day. Calibration analysis of ADCP backscatter versus 70-kHz Simrad backscatter shows a positive functional relationship, with 4 units of relative acoustic backscatter intensity equal to 1 dB of Sv.

2:45

1pAB6. Bottom-mounted hydrophones used to investigate sperm whale click and swim characteristics. Eva-Marie Nosal and L. Neil Frazer (Dept. of Geology and Geophys., Univ. of Hawaii, 1680 East-West Rd., Honolulu, HI 96822)

A method that relies solely on the time difference between direct and reflected arrivals (DRTD) on individual hydrophones was previously used to localize a single sperm whale [E.-M. Nosal and L. N. Frazer, Appl. Acoust., Special Ed.: Detection and Localization of Marine Mammals using Passive Acoustics (in press)]. Regular clicks in recordings from five bottom-mounted hydrophones in the AUTEC range were used in the localization. In this study, the DRTD method is extended by incorporating time differences (between hydrophones) of direct arrivals to refine the track of the whale and to determine the time of generation of individual clicks. Clicks, which have a multi-pulse structure, are analyzed in detail to determine changes in interpulse intervals (IPI) with time. The resulting time, location, and IPI information are used to investigate swim patterns and directionality characteristics of regular sperm whale clicks. [Work supported by ONR. Data provided by NUWC for the 2nd International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics.]

3:00

1pAB7. 3-D tracking of sperm whale dive profiles from a towed hydrophone array in ray-refractive environments. Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238, thode@mpl.ucsd.ed), Tom Norris (Sci. Applications Intl. Corp., San Diego, CA), Alyson Azzara, Ann Jochens, and Doug Biggs (Texas A&M Univ., College Station, TX 77843)

During the past 3 years the U.S.-based Sperm Whale Seismic Study (SWSS) has supported the development of a mobile three-dimensional passive acoustic tracking system. Acoustic data were recorded on a single four-element wide-aperture hydrophone array, towed at 80–100-m depth. The array was subdivided into forward and rear two-element subarrays, separated by 200-m aperture. To fix a location of an impulsive sound, three quantities are measured: the time difference between direct-path arrivals on the forward and rear subarrays, the time difference between the direct and surface-reflected paths on the rear subarray, and the acoustic bearing measured on the rear subarray. These quantities, combined with independent measurements of hydrophone depths and cable inclination, are postprocessed using a MATLAB GUI into range-depth position fixes by implementing an efficient numerical procedure that uses precomputed look-up tables to account for ray-refraction effects caused by depth-dependent sound-speed profiles. The predicted time difference between the direct and surface-reflected paths on the forward subarray is compared with the measured data to adjust for array tilt and depth measurement error. A by-product of this processing is a semiautomated detector to detect distinctive creak sounds. [Work supported by the U.S. Minerals Management Service and the Industry Research Funders Coalition.]

3:15

1pAB8. Direct underwater visual and acoustic observations of echolocation behavior of sperm whales around a longline. Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla CA 92093-0238), Jan Straley, Kendall Folkert, Victoria O'Connell (Univ. of Alaska Southeast, Sitka, AK 99835), Linda Behnken (Alaska Longline Fishermens Assoc., Sitka, AK 99835), and Amanda Koltz (Cornell Univ., Ithaca, NY 14853)

Sperm whales (*Physeter macrocephalus*) have learned to approach and deprecate fishing vessels longlining for black cod in the Gulf of Alaska. In May 2006, off Sitka, Alaska, a video camera was placed inside an underwater housing with an external hydrophone mount, attached to a longline that had been partially hauled to the surface, and lowered between depths of 90 and 120 m, with black cod (*Anoploploma fimbria*) attached 2 to 4 m above the camera as bait. The camera was mounted facing towards the surface to exploit the ambient light and to eliminate the need for artificial lighting. An additional hydrophone was deployed at 17-m depth during the 40–60-m deployments, and a third autonomous acoustic recorder was usually mounted 1.3 m below the camera. During one encounter a whale investigated the line, producing characteristic creak sounds that were recorded on the three hydrophones, and which were subsequently time-aligned using vessel noise as a reference, in order to permit acoustic tracking. During the second encounter a whale interacted with two fish within 4 m of the camera, but only the camera hydrophone was available. Preliminary analysis of the echolocation behavior displayed by both animals is presented. [Work sponsored by the National Geographic Society and the North Pacific Research Board.]

3:30−3:45   Break

3:45

1pAB9. Three-dimensional passive acoustic tracking of sperm whale behavior with widely separated vertical arrays in the Gulf of Alaska. Christopher O. Tiemann (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713), Aaron M. Thode (UCSD, La Jolla CA 92093-0238), Jan Straley, Kendall Folkert (Univ. of Alaska Southeast, Sitka, AK 99835), and Victoria O'Connell (Alaska State Fish and Game, Sitka, AK 99835)

The Southeast Alaska Sperm Whale Avoidance Project (SEASWAP) was started in 2002 to study the increasing depredation encounters between sperm whales and longline fishermen near Sitka, Alaska. In 2005, the acoustic component of the study included passive acoustic data from four autonomous recorders attached to the anchor lines of two longlines spaced 3 miles apart, essentially turning them into temporary vertical arrays and allowing the monitoring of sperm whale vocal activity during fishing operations. Although the use of multipath arrival information from recorded sperm whale clicks has been a standard procedure for range-depth tracking, with the added knowledge of azimuthally dependent bathymetry relative to an array, a three-dimensional track of whale motion can be obtained from the same multipath information. The evolution of multipath arrival patterns is matched to range-, depth-, and azimuth-dependent modeled arrival patterns to generate an estimate of whale motion. The use of an acoustic propagation model can also account for all waveguide physics such as interaction with range-dependent bathymetry and ray refraction. Acoustic tracking results concurrent with whale encounters will be presented and compared to visual sightings.

4:00

1pAB10. Variation in sound production in relation to behavior of bottlenose dolphins (*Tursiops aduncus*). Michelle Lemon (Marine Mammal Res. Group, GSE, Macquarie Univ., Sydney, Australia, Defence Sci. and Tech Org., Pyrmont, Australia, and michelle.lemon@dsto.defence.gov.au), Douglas H. Cato (Defence Sci. and Tech Org., Pyrmont, New South Wales, Australia), Tim P. Lynch (Jervis Bay Marine Park, Huskisson, New South Wales, Australia), and Robert G. Harcourt (Macquarie Univ., Sydney, Australia)

Dolphins are social animals that inhabit an opaque environment where visual cues are often limited, thus acoustic signals play an important role in inter- and intragroup communication. Phonations may be used to signal conspecifics, in navigation, or in the detection of prey, and may vary with behavior, environmental conditions, or geographic location. Variation in the acoustic and surface behavior of two geographically separated and genetically differentiated populations of bottlenose dolphins in Southeastern Australia were investigated over a 3-year period. Two hypotheses were tested: (1) vocal repertoire varies with behavior; and (2) variation in vocal repertoire may have arisen due to geographic isolation. Overall rate of sound production by the dolphins was similar in each field site but varied significantly with behavioral activity. Call rates were highest with behavior that involved increased arousal, such as socializing and foraging. Whistle rates and the duration of echolocation click bouts were similar for each location, however, burst-pulsed sounds varied significantly with behavioral state and between the field areas. This difference may relate to differences in social behavior between populations and be driven by variability in social structure and environmental conditions. [Research supported by Australian Geographic, PADI Foundation and Macquarie University.]

4:15

1pAB11. Observation studies based on echolocation click signals of Irrawaddy dolphins in Chilika lagoon. Rajendar Bahl (Ctr. for Appl. Res. in Electron., IIT Delhi, Hauz Khas, New Delhi, India, rbahl@care.iitd.ernet.in), Tamaki Ura, Tomoki Inoue, Harumi Sugimatsu, Takeshi Sakamaki (Univ of Tokyo, Meguro, Tokyo 153-8505, Japan), Junichi Kojima (KDDI R&D Labs. Inc., Saitama 356-8502, Japan), Tomonari Akamatsu (Natl. Res. Inst. of Fisheries Eng., Ibaraki 314-0408, Japan), Hideyuki Takahashi (System Giken Co. Ltd., Kanagawa 253-0085, Japan), Ajit Pattnaik, Muntaz Khan (Chilika Dev. Authority, Orissa, India), Sandeep Behera (Freshwater and Wetlands Prog., New Delhi 110003, India), and Sudhakar Kar (Off. of Princip. CCF (Wildlife) Chief Wildlife Warden, Bhubaneswar 751012, India)

A resident population of Irrawaddy dolphins, *Orcaella brevirostris*, in Chilika lagoon in eastern India has been recently monitored based on their echolocation clicks. The free-ranging animals were monitored over several days in their natural habitat. A horizontal three-hydrophone line array was used for click data collection and off-line analysis of click trains to obtain underwater movements in shallow water of typical depth less than 3 m. The lagoon waters being placid provided a good surface echo that has been used to obtain depth information of the individual click sounds. This paper augments previous data on the broadband click characteristics [S. M. Van Parijs, G. J. Parra, and P. J. Corkeron, J. Acoust. Soc. Am. **108**, 1938–1940 (2000)] and also provides a glimpse into the individual behavior of the dolphins. An estimate of the click lag time of the animal's biosonar has also been obtained for the first time for a free-ranging animal.

4:30

1pAB12. On the use of the Teager energy operator for the detection of clicks in free-ranging dolphins. Marie A. Roch (Dept. of Comput. Sci., 5500 Campanile Dr., San Diego State Univ., San Diego, CA 92182-7720), Melissa S. Soldevilla, Jessica C. Burtenshaw, E. Elizabeth Henderson, Sean M. Wiggins, and John A. Hildebrand (Univ. of California at San Diego, La Jolla, CA 92093-0205)

The detection of individual clicks produced by odontocetes is useful for many bioacoustic tasks such as the analysis of interclick interval, classification of odontocete species, etc. Automation of this task is required for

the large data sets gathered by remote instruments that can collect terabytes of information over short time periods. A method is presented that locates individual clicks using the Teager energy operator, which is a time-domain metric designed to measure fast variations in signal energy. The Teager energy of a high-pass filtered signal is analyzed by a rule-based classifier that determines the location of individual clicks. The click start can be determined with very high accuracy. Click ends are reasonably well characterized, but the ends of clicks are difficult to determine, even by human experts, due to the influence of external factors such as acoustic reflections. The proposed method is compared with existing techniques and is demonstrated on the calls of several species of free-ranging odontocetes resident in the Southern California Bight and the Gulf of Mexico: bottlenose dolphins (*Tursiops truncatus*), short and long beaked common dolphins (*Delphinus delphis*, *Delphinus capensis*), Pacific white-sided dolphins (*Lagenorhynchus obliquidens*), and Risso's dolphins (*Grampus griseus*).

4:45

1pAB13. The acoustic recording of bottlenose dolphin (*Tursiops truncatus*) phonations: Problems and contributions. Carmen Bazúa-Durán and Axel Herrera-Hernández (Lab. Acústica, Facultad de Ciencias, UNAM, Ciudad Universitaria, 04510 México, D.F., México, bazua@servidor.unam.mx)

Bottlenose dolphin sounds have been studied for decades; however, more needs to be investigated. This work demonstrates that results obtained in wild dolphin studies greatly depend on the recording system used. Dolphin phonations were simultaneously recorded with two systems, a narrow-band (<24 kHz) and a broadband (<180 kHz), and were analyzed using spectral characteristics (narrow-band: 512 points, Hamming window, 50% overlap; broadband: 2048 points, Hamming window, 50% overlap) to determine the acoustic repertoire as echolocation click trains (EC), whistles (W), and burst pulses (BP). As expected, the number of BP and EC recorded was considerably larger with the broadband system than with the narrowband; whereas the number of W recorded was greater with the narrowband system than with the broadband (it was expected to obtain the same number of W with both systems). As W are less intense than BP or EC, when gain is adjusted for recording BP or EC, the recording system is not able to record the less intense W. Therefore, it is not possible to capture the same number of dolphin phonations with the same recording system. One should use at least two different recording systems to adequately describe dolphin phonations. [Work supported by CONACyT-Campeche and PAPIIT-UNAM.]

5:00

1pAB14. Acoustic monitoring of resident, offshore, and transient killer whales off the Washington coast. Erin M. Oleson, Sean M. Wiggins, and John A. Hildebrand (Scripps Inst. of Oceanogr., Univ. of California San Diego, 9500 Gilman Dr., #0205, La Jolla, CA 92093, eoleson@ucsd.edu)

Three acoustically distinct populations of killer whales representing each of the known ecotypes (resident, offshore, transient) were recorded in the Summer–Fall of 2004 off the southern Olympic Coast of Washington. Two high-frequency acoustic recording packages (HARPs) continuously recording at 80-kHz sample rate were deployed to assess the seasonal occurrence of vocal odontocetes in this region. From mid-July to early-October the population-specific discrete calls of killer whales were heard on 8 days and were classified to population by Volker Deecke (UBC) and John Ford (DFO-Canada) using an acoustic ID catalogue. West Coast Transient killer whales producing calls of the California dialect were heard on three occasions from August through October. Offshore killer whales were heard twice in August–September, and Northern Resident killer whales were heard once in August. Although Northern Resident killer whales have been extensively studied within Puget Sound and coastal British Columbia, they have been visually sighted only once off the northern Olympic Peninsula, making their detection at this offshore southerly location unique. Endangered Southern Resident killer whales were not heard at this site from July–October. Analysis of year-round data from a site further offshore is underway. [Funded by Chief of Naval Operations-N45.]

5:15

1pAB15. High-frequency Acoustic Recording Package (HARP) for long-term monitoring of marine mammals. Sean M. Wiggins, Chris Garsha, Greg Campbell, and John A. Hildebrand (Univ. California-San Diego, 9500 Gilman Dr., 0205, La Jolla, CA 92093-0205, swiggins@ucsd.edu)

Advancements in low-power and high-data-capacity computer technology during the past decade have been adapted to autonomously record sounds from marine mammals over long time periods. Acoustic monitoring has advantages over traditional visual surveys including greater detection ranges, continuous long-term monitoring in remote locations under various weather conditions, and lower cost. However, until recently, the technology required to autonomously record whale sounds over long durations has been limited to low-frequency (' 1000 Hz) baleen whales. The need for a broader-band, higher-data capacity system capable of autonomously recording toothed whales and other marine mammals for long periods has prompted the development of a High-frequency Acoustic Recording Package (HARP) capable of sample rates up to 200 kHz. The HARP design is described and data analysis strategies are discussed using examples of broadband recorded data. [Work supported by the U.S. Navy CNO-N45.]

Session 1pAOa

Acoustical Oceanography and Underwater Acoustics: Acoustic Scattering by Aquatic Organisms II

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Masahiko Furusawa, Cochair
Tokyo Univ. of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo, 108-8477, Japan

Chair's Introduction—1:00

Contributed Papers

1:05

1pAOa1. Broadband backscatter measurements of individual Hawaiian mesopelagic boundary community animals. Whitlow W. L. Au (Hawaii Instititute of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 86734) and Kelly J. Benoit-Bird (Oregon State Univ., Corvallis, OR 97331-5503)

Broadband simulated dolphin echolocation signals were used to measure the *ex situ* backscatter properties of mesopelagic boundary community animals. The signal had a peak frequency of 100 kHz, 3-dB bandwidth of 35 kHz, and duration of 80μs. Subjects were captured by trawling with a 2-m-opening Isaacs-Kidd midwater trawl. Backscatter measurements were conducted on the deck of the ship in a 2000-L seawater plastic tank with the transducer placed on the bottom, pointed upwards. Anesthetized animals were suspended in the transducer's beam by nylon monofilament lines connected to a support device constructed out of dowels. Backscatter measurements were obtained in both the dorsal and lateral aspects for seven myctophids and only in the dorsal aspect for the other 13 myctophids, 6 shrimps, and 3 squids. The echoes from the myctophids and shrimps usually had two highlights, one from the surface of the animal nearest the transducer and a second probably from the signal propagating through the body of the subject and reflecting off the opposite surface of the animal. The squid echoes consisted mainly of a single highlight but sometimes with a low-amplitude secondary highlight. The target strength had a rippled pattern in the frequency domain showing a distinct null.

1:20

1pAOa2. Acoustic measurements of volume backscattering strength from red tide phytoplankton, *Cochlodinium polykrikoides*. Tae-Hoon Bok, Dong-Guk Paeng (Faculty of Marine Industry Eng., College of Ocean, Cheju Nat'l Univ., Ara-dong, jeju-si, jeju-do, Korea, 690-756), Eunhye Kim, and Jungyul Na (Hanyang Univ., Sa-dong, Ansan-si, Kyeonggi-do, Korea, 425-791)

This research is to measure the volume backscattering strength, Sv, of the *Cochlodinium polykrikoides* (the main species of red tide in Korea) in order to investigate the possibilities of early detection of red tide using high-frequency transducers (5, 10 MHz). First, the Sv from the cultured *C. polykrikoides* was measured with the population density in the laboratory. Next, ultrasonic transducers were attached to the side of a research vessel, and the Sv was measured over 9-km ship track near Yeosu in the Korean southern sea. Seawater at three stations was sampled, and the number of *C. polykrikoides* was counted to compare the theoretical Sv with the measured one. The results show that the measured Sv was in agreement with the theoretical one. The Sv when the light was on was measured to be higher than that when the light was off in the laboratory, probably due to the increase of size and density of *C. polykrikoides* during photosynthesis. In the field experiment, the Sv was measured in real time with the position information from the Global Positioning System data. This research proposes the use of underwater ultrasonic technology for early detection of red tide in real time.

1:35

1pAOa3. Measurements of target strength spectra of metal spheres and fish by using broadband sonar signals of dolphin. Tomohito Imaizumi, Masahiko Furusawa (Tokyo Univ. of Marine Sci. and Technol., 4-5-7 Kounan, Minato Tokyo, 108-8477, Japan, imaizumitomohito@yahoo.co.jp), Tomonari Akamatsu (Fisheries Res. Agency of Japan, Kamisu, Ibaraki 314-0408, Japan), Yasushi Nishimori, and Shinji Ogawa (Furuno Electric Co., Ltd., Nishinomiya-city, 662-8580, Japan)

Dolphins identify prey by using their broadband sonar signals. Considering that the spectrum of fish target strength (TS) leads to identification of objects, we measured the spectrum using dolphin's sonar signals with a broadband system. The TS spectrum is obtained by dividing the echo spectrum by the incident spectrum while setting the projector and hydrophone face to face. First, we measured the broadband form function of a tungsten carbide and copper sphere in a water tank and confirmed good agreements with theoretical value. Second, we measured the spectra of TS of three species of anaesthetized fish in a water tank. Although the spectra showed nearly similar tendencies of past measurements, they changed according to the species, the individual, and the tilt angle. Third, we measured TS spectra of fish suspended by nylon monofilament line in the open sea in semi-*in-situ*. We tried to remove the directivity of transducers by a tungsten carbide sphere suspended near the fish. The change of the TS spectra should be important information for the fish identification.

1:50

1pAOa4. Variations of echosounder transducer performance versus water temperature. David A. Demer and Josiah Renfree (NOAA/ NMFS/Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA, 92037)

Electroacoustic transducers are central components of multi-frequency echosounders used in remote target identification and acoustical surveys for fish and zooplankton. While these transducers, constructed from multiple materials such as piezoelectric ceramics, polyurethanes, and metals, are designed to be insensitive to performance changes versus time, pressure, and temperature, it has been shown that appreciable changes in echosounder system gains result from shifts in transducer resonant frequency, quality factor, and electrical impedance versus water temperature. Because it is standard practice to calibrate echosounder systems in one environment and apply the resulting gains to interpret data collected over a range of sea temperatures, the survey results may be significantly biased. The bias may be different for estimates derived from each echosounder frequency. The magnitudes of these frequency-dependent biases depend upon the temperature-dependent performances of the survey transducers, and the

range of temperatures encountered during the survey. Here, the performance has been measured for ten commonly used survey transducers versus water temperature ranging from 1 to 18 °C, using four different techniques. Temperature-dependent system gain corrections are proposed that can be used to minimize bias in multifrequency target identifications and biomass estimations. [Funding was provided by the NMFS Advanced Sampling Technology Working Group.]

2:05

1pAOa5. Calibrating split beam transducers at depth. Egil Ona and Geir Pedersen (Inst. of Marine Res., P.O. Box 1870, 5817, Bergen, Norway)

A new echo sounder probe for target strength measurements in deep water have been used to calibrate the depth performance of several pressure-stabilized split beam transducers. The probe carries three Simrad EK60 echo sounders and can be operated on 6500 m optical cable to about 1500-m depth. The transducers are mounted on a motorized, revolving platform. Digital angle sensors monitor transducer orientation, and an electronic compass also monitors transducer rotation. Calibration includes standard sensitivity parameters for split beam transducers, but also the equivalent beam angle and angle sensitivity at different depths. Examples of target strength data on marine animals will also be shown.

2:20

1pAOa6. Standard-target calibration of active sonars used to quantify aquatic organisms. Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The span of sonars being calibrated by the standard-target method continues to grow. In this retrospective of the past 25 years, the essence of the method is reviewed for both narrow-band and broadband systems. Applications to scientific echo sounders with operating frequencies from 18 kHz to 3.2 MHz are summarized. Applications to multibeam sonars that provide the water-column signal, with frequencies of 90, 200, 240, and 400 kHz, are also summarized. A planned application to a parametric sonar, with primary frequencies in the range 15–21 kHz and difference-frequency band 0.5–6 kHz, is described. The method is applicable to ordinary midfrequency sonars. In each example, particular standard targets are specified. Diameters vary from 10 mm for a sphere of tungsten carbide with 6% cobalt binder for use at low megahertz frequencies, to 60 mm for a sphere of electrolytic-grade copper for use at 38 kHz, to 280 mm for a sphere of aluminum alloy for use at low-kilohertz frequencies.

2:35

1pAOa7. New scientific multibeam systems (ME70 and MS70) for fishery research applications. Lars Nonboe Andersen, Sverre Berg, Ole Bernt Gammelster, and Even Borten Lunde (Simrad AS, P.O. Box 111, 3191 Horten, Norway, fish_research@simrad.com)

Fishery scientists have for many years been requesting calibrated multibeam systems specially designed for fishery research applications. Simrad AS, Norway, has, in cooperation with IFREMER, France, developed a new multibeam echo sounder (ME70) and with IMR, Norway, a new multibeam sonar (MS70) for fishery research applications. Both systems have 800 transmitting and receiving channels and are based on similar hardware and software platforms. The systems are characterized by calibrated narrow beams, low sidelobe levels, high instantaneous dynamic range, and operate in the frequency range 70–120 kHz. The echo sounder is designed for high operating flexibility, with a configurable fan of 1–45 roll/pitch stabilized split beams with opening angles ~2 to ~7. In addition, two configurable reference beams are available for comparison with standard echo sounder systems. The sonar is mounted on a drop keel, looking horizontally, covering a horizontal sector of 60 and a vertical sector of 45. This coverage matrix is generated using 500 roll stabilized beams, 25 beams horizontally with a resolution of ~3, and 20 beams vertically with a resolution of ~4. Both systems are designed for fish-stock assessment and fish-behavior studies. Results from system measurements will be presented.

2:50–3:10 Break

Invited Papers

3:10

1pAOa8. First results with the new scientific multibeam echo-sounder ME70. Verena Trenkel (Ifremer Ctr. de Nantes, BP 21105 44311, Nantes Cedex 03, France), Valerie Mazauric, and Laurent Berger (Ifremer Ctr. de Brest, Plouzane, France)

Multibeam echo-sounders and sonars are increasingly used in fisheries acoustics for studying school structure and behavior and for abundance estimation. However, data quality and a poor signal-to-noise ratio obtained with the existing multibeam echo-sounders preclude quantitative biomass estimation as carried out with the traditional single beam echo-sounders. In order to overcome these shortcomings, Ifremer and Simrad AS started in 2003 the development of the new and highly flexible calibrated multibeam echo-sounder ME70 designed for studying a wide range of scientific questions. This paper will present the full functionalities of the ME70 and preliminary results obtained with the system installed onboard the research vessel *Thalassa*. It will also provide first indications of its performance. Data have been collected on small pelagic fish species in the bay of Biscay, using simultaneously the traditional single beam echo-sounders and the ME70 with different beam configurations. Preliminary comparisons of fish school descriptors and biomass estimates based on single- and multibeam data allow us to quantify the effect of the gain in sampling coverage and the high angular resolution of the new system.

3:30

1pAOa9. First data from sea trials with the new MS70 multibeam sonar. Egil Ona, John Dalen, Hans Petter Knudsen, Ruben Patel (Inst. of Marine Res., P.O. Box 1870, 5817 Bergen, Norway), Lars Nonboe Andersen, and Sverre Berg (Simrad A/S, 3191, Horten, Norway)

The new multi-beam sonar, MS70, is a horizontally observing sonar yielding very high spatial resolution when operating all 500 beams, covering the frequency band 75–112 kHz. The sonar has undergone sea trials from *R/V G. O. Sars* from December 2005. The presentation includes preliminary results obtained during this period. Topics covered will be sonar performance objectives, calibration methods, and preliminary results from calibration and school data acquisition. Examples of raw and processed 3-D and 4-D data from a stationary and surveying vessel on small sprat schools and large herring schools will be shown.

1pAOa10. Development of a quantitative scanning sonar system and sea trial for a fish school of herring. Yasushi Nishimori (Furuno Elec. Co., Ltd., 9-52 Ashihara-cho, Nishinomiya, 662-8580, Japan), Kohji Iida, Tohru Mukai (Hokkaido Univ., Hakodate, 041-8611, Japan), Yong Tang (Dalian Fisheries Univ., Dalian, 116023, China), Masaki Ogata, Kohzoh Tokuyama, and Emi Okazaki (Furuno Elec. Co., Ltd., Nishinomiya, 662-8580 Japan)

A quantitative sonar system enabling data recording, calibration, and numerical analysis of fish schools was developed based on the commercial scanning sonar. This system is equipped with a newly developed algorithm for estimation of the abundance of a fish school using three-dimensional echo integration. Estimation was made on the condition that the target strength of all fish is identical and known and the weight of a single fish is known. The field recording was performed to evaluate the algorithm. The school of spring spawning herring were scanned three-dimensionally and recorded using two different scanning methods. Two methods were applied in different distances between the school and vessel. The volume and three-dimensional echo integration for these two recorded dates were calculated and compared to each other. It was found that the variance of three-dimensional echo integration is smaller compared with the variance of volume. A compensation of abundance estimation using average target strength was tried and its effectivity was examined.

Contributed Papers

4:10

1pAOa11. Measurement of scattering directivity and behavior of fish schools using multibeam sonar. George Cutter, Jr. and David Demer (NOAA Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92038)

Fish dispersion and abundance are commonly estimated from volume backscattering coefficients (sv) of animals located directly beneath the survey vessel, inside a narrow transducer beamwidth. Fish density (#/m^2) is estimated from a ratio of integrated sv and a representative backscattering cross-sectional area. Fish biomass is estimated from the product of fish density and the area sampled. Uncertainty in biomass estimates can arise from small sampling volumes, and the scattering directivities of the fish coupled with their orientations relative to the acoustic beam. Multibeam sonars can be used to increase the observational volumes, but add the complexity of a wider range of incident angles. If fish size and packing density are constant throughout a school, then the variation in sv versus aspect angle can be attributed to fish orientation. With modeling and multibeam sonar measurements, the following null hypotheses are explored: (1) The variability of sv versus roll angle is nearly uniform; (2) For measurements made with beams at or near vertical, sv variability is primarily due to the pitch angle; (3) For measurements with nonvertical beams, such as the outer beams of multibeam echosounders, sv variability is primarily due to the yaw angle.

4:25

1pAOa12. Advances in the evaluation of acoustic beamformed backscatter data for fisheries applications. Chris Malzone (The RESON Group, 100 Lopez Rd. Goleta, CA 93117, cmalzone@reson.com), Matthew Wilson (SonarData Pty Ltd, Hobart TAS 7001 Australia), and Pat Iampietro (CSU Monterey Bay, Seaside, CA 93955)

Single-beam echosounders are well-established tools for fisheries research. However, single-beam echosounder geometry is limited to a very small ensonification volume located within a 3 to 15-deg cone directly beneath the vessel. Recent advances in multibeam echosounder technology provide the option to collect backscatter and water-column data during an IHO order 1 hydrographic survey. Such comprehensive data sets are valuable for scientific and management purposes since they expand the volume in which water-column data are collected, and they provide as well a potential method to link the biology to the habitat. However, water-column information collected with multibeam echosounders poses several challenges, including increased data storage requirements, a need for data decimation, and survey methodology considerations. In this paper, concurrently acquired backscatter data from a split-beam scientific echosounder and multibeam echosounder are analyzed for relative abundance, spatial distribution, and avoidance behavior of pelagic organisms, primarily the market squid *Loligo opalescens* and various finfish. Aspects of the data sets are then explored, compared, and contrasted. Proposed solutions to some of the challenges posed by multibeam sonar water-column analysis are then reviewed and placed within a overall framework for fisheries applications.

4:40

1pAOa13. Mapping fish schools with a high-resolution multibeam sonar. Mark Trevorrow (Defence R&D Canada Atlantic, 9 Grove St., Dartmouth, NS, B2Y 3Z7 Canada, mark.trevorrow@drdc-rddc.gc.ca)

In late 2002 an evaluation of a high-resolution multibeam sonar (RESON SeaBat 8125) for mapping mid-water fish schools was performed in Knight Inlet, BC. The ship-mounted multibeam sonar operated at 455 kHz and provided 240 one-half-degree beams in the across-track plane with one-degree along-track beam width. Simultaneous measurements were made with a conventional 200-kHz echo-sounder. Acoustic calibrations of this multibeam sonar will be discussed, showing some range- and angle-dependent variations in sensitivity. The multibeam sonar was able to provide well-resolved mid-water images of fish schools (likely juvenile herring), sampling many fish schools missed by the echo-sounder and revealing cross-track fish school shape. Comparison of near-vertical volumetric scattering strengths between the multi-beam sonar and echo-sounder show reasonable agreement. Discussions on the combined use of conventional and multibeam sonars for fish surveys will be given.

4:55

1pAOa14. Multibeam assessments of Australian seabed habitats. Iain Parnum, Alexander Gavrilov, Paulus Siwabessy, and John Penrose (Ctr. for Marine Sci. and Technol., Curtin Univ. of Technol., G.P.O. Box UU1987, Perth, Western Australia 6845)

In the period 2003–2006 a major program of shallow seabed habitat assessment has been undertaken in Australian coastal seas. As part of the Coastal Water Habitat Mapping (CWHM) project of the national Cooperative Research Centre for Coastal Zone, Estuary and Waterway Management, research into optimal use of acoustic techniques for habitat assessment has been undertaken. Echo sounder, side scan, and multibeam acoustic systems have been used and evaluated in this program. Multibeam systems have been primarily employed with particular attention being given to combining topography data with associated measures of backscatter intensity. The emphasis on shallow waters, generally 50-m depth or less, and high-frequency multibeam systems with up to 240 beams has provided data sets with relatively high levels of spatial resolution. Measures of backscatter variation with incident angle have been incorporated into the interpretation process and a new angular dependence correction algorithm developed for production of backscatter imagery. The processing techniques developed in the CWHM project have been evaluated using towed video data for a variety of tropical and temperate coastal environments around Australia.

1pAOa15. Acoustic identification and density estimate of *Diaphus theta* (myctophidae) off Hokkaido, Japan. Hiroki Yasuma, Kazushi Miyashita (Field Sci. Ctr. for the Northern Biosphere, Hokkaido Univ., 3-1-1, Minato, Hakodate, Hokkaido, 041-8611, Japan), and Orio Yamamura (Hokkaido Natl. Fisheries Res. Inst., Kushiro, 085-0802, Japan)

Diaphus theta is the most abundant micronektonic fish in the northwestern Pacific. Field acoustic data and biological samples were obtained both day and night off of eastern Hokkaido, Japan in September 2006 to estimate the fish density and diel changes of its vertical distribution. Differences of acoustical target strength (TS) values between at 38 and 120 kHz were applied to identify *D. theta* echos using field echo data. Theoretical estimation using swimbladder acoustic scattering models showed that the TS difference (TS 120-38 kHz) of larger fishes (over 60 mm) was between −4 and 1 dB; that for smaller fishes (less 60 mm) was less than .4 dB. These values differed from those of other major component species of the deep-scattering layer, such as krill and pollock, suggesting that the echo from *D. theta* is acoustically identifiable. Diel vertical distributions of *D. theta* were estimated after identification. Relatively dense schools were observed at around 400-m depth during daytime, although the schools were scattered widely above 100-m depth during nighttime. These results matched those obtained by biological sampling. In the surface layer (above 100 m), the estimated fish densities were about tenfold higher at nighttime than in daytime.

1pAOa16. Measurement of fish school abundances in shallow sea using omnidirectional multibeam sonar. Yong Tang (Dalian Fisheries Univ., 52 Heishijiao-jie Shahekou, Dalian, 116023, China;tang@dlfu.edu.cn), Kohji Iida, Tohru Mukai (Hokkaido Univ., Hakodate, 041-8611, Japan), and Yasushi Nishimori (Furuno Elec. Co., Ltd., Nishinomiya, 662-8580 Japan)

Previous studies had been constructed in the use of commercial scanning sonar to estimate the abundance of pelagic fish school. However, a relatively narrow dynamic range and lack of linear signal processing limited the application of sonar. In this research, quantitative omnidirectional multi-beam sonar (FSV30 research version, 24 kHz, Furuno) was used to investigate a fish school in a shallow area. The raw volume backscattering strength signal (RSV, output signal with $20 \log r$ TVG and coefficient of multiple echo) of the fish school echoes were recorded and then its biomass was confirmed using purse seine capture. In postprocessing, the fish school echoes were extracted from the sonar echogram using the image processing method (polygon and a low-pass filter). The relationships between a school's biomass, its echo area, the maximum, and the average RSV of each ping were considered for various species. Results showed a positive correlation between the sonar-measured school area and school biomass, but no correlation between the RSV and school biomass. Also, the echo integration method for sonar survey using RSV in broad area and methods for averaging the target strength (TS) in three dimensions with respect to echo integration method were also discussed.

TUESDAY AFTERNOON, 28 NOVEMBER 2006   KAUAI ROOM, 1:00 TO 5:20 P.M.

Session 1pAOb

Acoustical Oceanography and Signal Processing in Acoustics: Acoustic Tomography for Coastal and Deep Water Applications II

James A. Mercer, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Arata Kaneko, Cochair
Hiroshima Univ., Graduate School of Engineering, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8527, Japan

Chair's Introduction—1:00

Invited Papers

1:05

1pAOb1. Precise measurement of travel time difference for reciprocal transmission of Central Equatorial Pacific tomography. Hiroyuki Hachiya (Dept. of Information & Image Sci., Chiba Univ., 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan), Toshiaki Nakamura (Natl. Defense Acad., Kanagawa 239-8686, Japan), Hidetoshi Fujimori, and Iwao Nakano (Japan Agency for Marine-Earth Sci. and Technol., Yokosuka 237-0061, Japan)

We present the estimation results of the travel time difference in reciprocal transmission data collected during the tomography experiment performed by Japan Agency for Marine-Earth Science and Technology (JAMSTEC) in the Central Equatorial Pacific. A precise estimation of travel time is essential for measuring ocean current using ocean acoustic tomography. We estimate the variation of the correlation of reciprocal propagated sound transmissions. Generally, reciprocal propagation sound waveforms that are transmitted at the same time are similar mutually, and it is considered that their correlation is high. However, it is found that reciprocal sound waveforms sometimes become different in the experiment. The vertical arriving angle of ray is measured using the phase differences between the receivers of a 30-m-long receiving array. The theoretical analysis of phase deviation using the Rice distribution is presented. From the ray identification results, the travel time difference between corresponding rays of reciprocal transmissions can be estimated. Since the phase of a ray signal is very stable, the precision of the travel time difference using phase difference is 6 times higher than the conventional estimation. The combination of the amplitude and phase information is effective for observing the ocean structure change.

1pAOb2. Ocean acoustic thermometry and the seasonal cycle of temperature in the North Pacific Ocean. Brian D. Dushaw, R. Andrew, B. Howe, J. Mercer (Appl. Phys. Lab., Univ. of Washington, 1013 N.E. 40th St., Seattle, WA 98105-6698, dushaw@apl.washington.edu), B. Cornuelle, M. Dzieciuch, W. Munk, P. Worcester (Scripps Inst. of Oceanogr., La Jolla, CA), T. Birdsall, K. Metzger (Univ. of Michigan, Ann Arbor, MI), D. Menemenlis, and R. Spindel (California Insitute of Technol., Pasadena, CA)

With several years of long-range (several Mm) acoustic propagation data obtained during the Acoustic Thermometry of Ocean Climate (ATOC) and North Pacific Acoustic Laboratory (NPAL) projects, the seasonal cycle of ocean temperature in the North Pacific can be examined. Acoustic transmissions have been made from a source located off the northern Californian coast and from a source located north of Kauai, HI to several receivers of opportunity located in the North Pacific Basin. The acoustic data are a high signal-to-noise measure of large-scale temperature with excellent temporal resolution. Although only a few realizations of the seasonal cycle are available, it is clear that inter- and intraannual variabilities are large contributions to the time series, with signal amplitudes comparable to the seasonal cycle. Such variabilities are likely advective in origin. The time scales for some of the changes in temperature are short, sometimes of order weeks. Not all available acoustic paths are suitable for assessing the seasonal cycle, however. Near Hawaii, the acoustic sampling does not extend to the near-surface waters, so seasonal variations there are not measured. The acoustic results will be compared to measures of the seasonal cycle by satellite altimetry, profiling floats, and the ECCO numerical ocean model.

1pAOb3. The Long-range Ocean Acoustic Propagation EXperiment (LOAPEX): An overview. James A. Mercer, Bruce M. Howe, Rex K. Andrew, Michael A. Wolfson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, mercer@apl.washington.edu), Peter F. Worcester, Matthew A. Dzieciuch (Univ. of California, La Jolla, CA 92093), and John A. Colosi (Naval Postgrad. School, Monterey, CA 93943)

The Long-range Ocean Acoustic Propagation EXperiment (LOAPEX), conducted in the NE Pacific Ocean, provided acoustic transmissions from a ship-suspended source at eight widely separated stations, and from a cabled acoustic source near the Island of Kauai, HI. The transmissions were received on several bottom-mounted horizontal hydrophone arrays distributed about the NE Pacific Ocean Basin and two, nearly colocated, vertical hydrophone line arrays spanning roughly 3500 m of the water column. Ranges varied from 50 km to several Mm. The goals of the experiment are (i) to study the evolution, with distance (range), of the acoustic arrival pattern and in particular the dependence of the spatial and temporal coherence; (ii) to investigate the nature of the deep caustics and the associated arrivals well below their turning depths; (iii) to analyze the effects of the ocean bottom near the bottom-mounted acoustic source cabled to Kauai; and (iv) to produce a thermal snapshot of the NE Pacific Ocean. The experiment goals, design, and methods are described as well as preliminary data results. [Work supported by ONR.]

1pAOb4. The SPICEX (Spice Experiment) component of the 2004 North Pacific Acoustic Laboratory (NPAL) experiment: An overview. Peter F. Worcester, Matthew A. Dzieciuch, Lora J. Van Uffelen, Daniel L. Rudnick, Bruce D. Cornuelle, and Walter H. Munk (Scripps Inst. of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093)

The objectives of the SPICEX component of the 2004 North Pacific Acoustic Laboratory (NPAL) experiment were (i) to elucidate the relative roles of internal waves, ocean spice, and internal tides in causing acoustic fluctuations; (ii) to understand the acoustic scattering far into the geometric shadow zone beneath caustics previously seen on bottom-mounted SOSUS receivers (shadow-zone arrivals); and (iii) to explore the range dependence of the fluctuation statistics. Two vertical line array receivers deployed in close proximity in the central North Pacific Ocean, together spanning much of the water column, recorded transmissions from broadband 250-Hz sources moored at the sound channel axis (750 m) and at surface conjugate depth (3000 m) at ranges of 500 and 1000 km from June through November 2004. Extensive environmental measurements were made along the acoustic path. The acoustic data from SPICEX reveal the vertical structure of the deep shadow zone arrivals. Internal waves are able to account for much of the observed vertical scattering in the reception finale, but other mechanisms seem necessary to account for the early shadow zone arrivals. Receptions from the abyssal sources show scattering occurring predominantly along the acoustic time fronts, as previously suggested by simulations, rather than across them.

1pAOb5. An overview of BASSEX (Basin Acoustic Seamount Scattering Experiment). Arthur B. Baggeroer (MIT, Cambridge, MA 02139), Kevin D. Heaney (OASIS, Inc., Lexington, MA), Peter F. Worcester, Matthew Dzieiuch (UCSD, La Jolla, CA 92037), James Mercer, Rex Andrew, and Bruce Howe (Univ. of Washington, Seattle, WA 98105)

BASSEX exploited the signals transmitted from the SPICEX, LOAPEX, and NPAL Kauai sources to examine the forward scattering from the Kermit-Roosevelt Seamounts near the Mendocino Fault in the northeast Pacific. The receiver was the FORA (four octave research array) from Pennsylvania State University. This array was towed at 300 m and had approximate resolutions of 2 and 6 deg for the 250- and 75-Hz signal transmissions. These were beamformed and pulse compressed for analyzing the multipath receptions from the sources. Source-receiver ranges covered from 250 to 1600 km provide a very rich data set. Data for forward- and backscattering from the slope near the NPAL source north of Kauai were also obtained. The analysis so far has demonstrated forward-scattering shadow convergence zones as well as diffraction from the seamounts. The upslope/downslope path structure near the Kauai source has also been analyzed. Both ray path and full wave acoustic modeling have been used for path identification as well as a theoretical model for modal scattering from a cone within an acoustic wave guide. [Work supported by ONR Code 321, Ocean Acoustics.]

1pAOb6. Ray and finite frequency kernel stability in the presence of realistic eddies. Bruce D. Cornuelle and Matthew A. Dzieciuch (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr, La Jolla CA 92093, bcornuelle@ucsd.edu)

Many recent studies have focused on the instability of ray paths (and therefore ray sampling kernels) in an ocean sound-speed field perturbed by internal waves. In many parts of the world's oceans there is synoptic scale eddy and planetary wave activity that produces sound-speed anomalies of larger scale than the internal waves, and of equal or larger magnitude. These are the features to be imaged by ocean tomography experiments, but the linearized forward problem that relates the travel time anomaly of the observed rays to the sound-speed anomalies may be invalid for realistic levels of eddy variability. This problem has been studied in the past for ray kernels by a number of authors, some of whom have concluded that nonlinearity requires iteration of the inversion process. We will compare and contrast the linearity of ray travel time kernel, finite frequency travel time kernels, and finite frequency kernels for other observables, such as pulse shape, in the presence of realistic variability for the N. Philippine sea, as estimated by ocean models and historical data. The travel time signals and nonlinearity due to the synoptic variability will be compared to that for internal waves. [Work supported by ONR.]

3:05–3:20 Break

Contributed Papers

1pAOb7. Seafloor hydrophone and vertical geophone observations during the North Pacific Acoustic Laboratory/Long-range Ocean Acoustic Propagation Experiment (NPAL/LOAPEX). Ralph A. Stephen (Woods Hole Oceanog. Inst., Woods Hole, MA 02543-1542, rstephen@whoi.edu), James A. Mercer, Rex K. Andrew (Univ. of Washington, Seattle, WA 98105-6698), and John A. Colosi (Naval Postgrad. School, Monterey, CA 93943)

Four OBSs (ocean bottom seismometers), consisting of one hydrophone and one vertical component geophone channel each, were deployed below the deep vertical line array (DVLA) on the North Pacific Acoustic Laboratory/Long-range Ocean Acoustic Propagation Experiment (NPAL/LOAPEX). The OBSs were deployed at 2 km offset north, south, east, and west from the nominal DVLA location. Data, sampled at 500 sps, were recorded continuously from the deployment in September 2004 for over 100 days. The OBSs were on the seafloor in about 5000 m of water and would be in the shadow zone for long-range, ducted propagation in the sound channel. The goals of the deployment were (1) to quantify the amount of energy that leaks out of the sound channel into the shadow zone; (2) to measure the relative sensitivity (signal-to-noise) of seafloor hydrophones and vertical component geophones to long-range signals; and (3) to study the physics of earthquake generated T-phases. Leakage was observed out to 3200 km on the vertical component geophone and out to 1000 km on the hydrophone. The ratio of the pressure to vertical velocity varies between arrivals on the same trace. [The OBS deployments were funded by contributions from ONR, NSF, and WHOI. Data analysis funded by ONR.]

1pAOb8. Mean acoustic field in long-range ocean acoustic propagation experiment (LOAPEX). Rex K. Andrew, Frank S. Henyey, James A. Mercer, Bruce M. Howe (Appl. Phys. Lab., Univ of Washington, 1013 NE 40th St., Seattle, WA 92105), Peter F. Worcester, and Matthew A. Dzieciuch (Univ. of California at San Diego, La Jolla, CA 92093)

We extract the low-frequency (70-Hz) mean acoustic pressure field, and the mean acoustic field up to time shifts, from the LOAPEX receptions on the vertical line arrays, for transmissions at various ranges. Means are taken over time intervals of order a day. Source and receiver motion is removed from the data, and possibly tides and other slow phenomena. The experimental results are compared to predictions from theoretical calculations assuming scattering by internal waves. The theoretical calculations make the Markov approximation that the sound-speed fluctuation correlations can be replaced by an operator at a single range. The calculations use modes rather than rays, because of the very low frequency, thus differing from the version of the Markov approximation that assumes delta-correlated sound-speed fluctuations.

1pAOb9. Mixed layer effects on acoustic arrival structure. Lora J. Van Uffelen, Peter F. Worcester, Matthew A. Dzieciuch, and Daniel L. Rudnick (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238, lvanuffe@ucsd.edu)

Oceanic mixed layers develop as the summer sun warms the upper layer of the water column and turbulence from early autumn storms stirs this heat downward, forming a well-mixed layer with constant temperature and salinity. The mixed layer is therefore thinnest in summer, when the surface is warm and the winds are relatively weak, and deepest in late winter. Receptions from moored 250-Hz acoustic sources at ranges of 500 and 1000 km in the North Pacific were recorded on vertical line array receivers from June through November 2004, spanning the early development of the winter mixed layer. The time-evolving character of the acoustic arrival structure throughout this transmission period will be investigated to explore the effect of a changing mixed layer environment on long-range acoustic propagation. Particular consideration will be given to variability in the early branches of the acoustic timefronts, which correspond to steep rays interacting with the upper ocean.

1pAOb10. Analysis of low-mode arrivals at 500 and 1000-km ranges in the North Pacific. Kathleen E. Wage (George Mason Univ., 4400 Univ. Dr., MSN 1G5, Fairfax, VA 22030, kwage@gmu.edu), Peter F. Worcester, and Matthew A. Dzieciuch (UCSD, La Jolla, CA 92093)

Experimental measurements of low-mode signals in the deep ocean are needed to verify statistical models and to design improved processing algorithms for these signals. Several long-range propagation experiments provided estimates of low mode statistics at ranges between 3 and 5 megameters [Wage *et al.*, J. Acoust. Soc. Am. **117**, 1565–1581 (2005)]. The 2004–2005 SPICE experiment included measurements at shorter ranges and a higher frequency than previous experiments. A 40-element vertical line array (VLA) received signals from two broadband sources at ranges of 500 and 1000 km. Both sources were moored at depths near the sound-channel axis; thus, they directly excited the lowest modes. The VLA was designed to spatially resolve 20 modes at the center frequency of 250 Hz. In addition to hydrophones the SPICE array included temperature and salinity sensors, which can be used to infer sound speed, and hence mode shapes, at the VLA. This talk describes an analysis of the low-mode arrivals in the SPICE data set. The presentation will include a discussion of mode filter design, focusing on how often the filter coefficients need to be updated to avoid problems with environmental mismatch. [Work supported by ONR.]

4:20

1pAOb11. Seamount geo-acoustic inversion from an acoustic scattering experiment (NPAL-04). Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039), Arthur B. Baggeroer (MIT, Cambridge, MA 02139), Peter F. Worcester, and Matthew A. Dzieciuch (Univ. of California, La Jolla, CA 92093)

It is thought that the deep ocean has thick, soft sediment due to millennia of build up from detritus. An open question is the sediment thickness on vast seamounts that rise many km from the ocean floor. Acoustic scattering from the Kermit-Roosevelt seamount (39.2O N 146.3 O W) in the central Pacific during the Basin Acoustic Seamount Scattering Experiment (BASSEX-04) in August of 2004 is used to address this question. The Kermit-Roosevelt seamount rises from a floor of 5400 m to a maximum height of 977 m below the surface. Broadband transmissions (100 Hz bandwidth, 250 Hz center frequency, 12.276 s period length) from two sources suspended (800 m depth, ranges of 521 and 625 km) were received on towed array (300 m depth, 200 m aperture) in the vicinity of the seamount. Combining array gain and time-bandwidth gain from the matched filter it is possible to clearly identify energy that has scattered from the seamount in 15 transmissions. These receptions are used to infer the sediment thickness and type as a function of the water depth. An efficient PE computational approach is applied as well as a reflection-coefficient ray-based approach.

4:35

1pAOb12. Mode processing of transient signals using a deficient vertical receiving array, with application to LOAPEX measurements. Ilya A. Udovydchenkov, Michael G. Brown (RSMAS, AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149), Peter F. Worcester, Matthew A. Dzieciuch, Lora J. Van Uffelen (Univ. of California at San Diego, La Jolla, CA, 92093), James A. Mercer, Bruce M. Howe, and Rex K. Andrew (Univ. of Washington, Seattle, WA 98195)

Mode processing of transient signals in a deep ocean environment using a deficient vertical receiving array is considered. Here, ''deficient'' implies that the vertical array is too sparse to resolve modes and/or has gaps. It is well known that mode processing of cw signals using a deficient array leads to significant modal cross-talk, i.e., nonzero projections of mode n energy onto mode m, $m \neq n$. It is shown, using simulated wave fields, that when phase-coherent processing is employed, transient signals are less sensitive to modal cross-talk than their cw counterparts. An explanation of this behavior is provided. Mode processing of a subset of the measurements made during the LOAPEX experiment at approximately 1 Mm range with $f_0 = 75$ Hz, $\Delta f = 35$ Hz is shown to give robust estimates of modal group time spreads for mode numbers $0 \leq m \leq 70$. [Work supported by ONR.]

4:50

1pAOb13. Horizontal coherence of tracked arrivals in the North Pacific Acoustic Laboratory98 (NPAL98). Matthew A. Dzieciuch (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA 92093, mad@ucsd.edu), Michael D. Vera (The Univ. of Southern Mississippi, Hattiesburg, MS), and The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester)

The NPAL98 experiment offered a unique opportunity to study horizontal coherence because of the deployment of a billboard array. The billboard consisted of five vertical line arrays with a 3600-m horizontal aperture listening to a 75-Hz acoustic signal 3500 km away. Original estimates of horizontal coherence were complicated by the bottom-interacting sound field, which made it difficult to track individual arrivals. Thus, previous studies of the horizontal coherence in this experiment were limited to the arrival pattern in aggregate. By using a turning-point filter and a rake correlator, it is now possible to track individual arrivals during the NPAL98 experiment. We can thus measure the coherence of individual early non-bottom-interacting arrivals. Results are compared to recent calculated predictions using both Monte Carlo and path-integral methods for the same environment. [Work supported by ONR.]

5:05

1pAOb14. The range evolution of the mean wavefront intensity for the long-range ocean acoustic propagation experiment (LOAPEX) off-axis source transmissions. Jinshan Xu (MIT/WHOI Joint Program in Oceanogr. and Oceanog. Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), John A. Colosi (Naval Postgrad. School, Monterey, CA 93943), James A. Mercer, Bruce M. Howe, Rex K. Andrew (Univ. of Washington, Seattle, WA 98105), Peter F. Worcester, and Matthew Dzieciuch (UCSD, La Jolla, CA 92093)

One of the main objectives of the NPAL 2004 experiment, LOAPEX (Long-range Ocean Acoustic Propagation EXperiment), which was conducted between 10 September and 10 October 2004, was to better understand the roles of scattering and diffraction in general. The LOAPEX measurement provided acoustic transmission data for ranges of 50, 250, 500, 1000, 1600, 2300, and 3200 km. By placing the source off-axis in order to avoid exciting low-order modes, we are able to study phenomena of the significant in-filling of acoustic energy into the finale region. Our focus will be on the transmissions for the off-axis source location (nominally 350-m depth), and the acoustic receptions as recorded on the 1400-m-long axial receiving array. The observation of the mean intensity of the wavefront arrival pattern at each range will be compared to deterministic ray and parabolic equation calculations. The following questions will be addressed here: (1) How does high angle acoustic energy from an off-axis source transfer energy to low angles in the axial region of the waveguide? (2) What are the relative contributions from diffraction and scattering? (3) How does this energy transfer scale with range? [Work supported by ONR.]

Session 1pBB

Biomedical Ultrasound/Bioresponse to Vibration and Signal Processing in Acoustics: Acoustic Tomography in Tissue

Jeffrey A. Ketterling, Cochair

Riverside Research Inst., 156 William St., New York, NY 10038-2609

Iwaki Akiyama, Cochair

Shonan Inst. of Technology, 1-1-25 Tsujido-nishikaigan, Fujisawa 251-8511, Japan

Chair's Introduction—1:35

Invited Papers

1:40

1pBB1. Acoustic computed tomography: Waves of approaches. James F. Greenleaf (Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55901)

Computed tomography (CT) using acoustic waves was invented at about the same time as MRI. The method of computing attenuation from projections was a direct analog of the x-ray CT approach. However, the method of computing speed from the time of flight of the acoustic ray was a larger logical jump. The field started with straight ray, earliest arrival time methods and has matured to what can be termed the full wave equation inversion method. The extension from two dimensions to three, although essential in the body, did not find a practical solution until very recently. The inverse problem represented by this field has found the attention of many investigators over the years. Practical medical implementations continue being investigated by several investigators more than 30 years after the first invention of the method. Acoustic tomography has been applied in nonmedical areas including oceanography, seismic, and solar sciences. Time awaits its success in medicine.

2:00

1pBB2. From laboratory to clinical trials: An odyssey of ultrasound inverse scattering imaging for breast cancer diagnosis. Steven A. Johnson, David T. Borup, James W. Wiskin, Michael Berggren, Barry Hanover, Frank Setinsek, Scott Olsen, and Karleen Callahan (Techniscan, Inc., 1011 East Murray Holladay Rd., Salt Lake City, UT 84117)

Techniscan has completed a local phase of clinical trials of a unique inverse scattering breast cancer scanner. The scanner is based on transmission of plane waves from 0.3 to 2 MHz through the breast to a receiver on the opposite side of the breast. The receiver consists of a 2-D array of six rows by 160 columns. These elements are multiplexed by a factor of 12 into 80 analog to digital converters. The dual transducer-head design allows the inversion of the transmission data to create two independent quantitative images, one of speed of sound and a second of attenuation. The two heads are mechanically attached to a hollow rotating shaft, to provide 360 deg of mechanical motion. Typically, 180 to 360 steps per 360-deg rotation and 2-mm vertical steps are used to create a data set that is inverted into a 3-D image of 240 by 240 by 70 voxels. A unique set of custom circuit boards was constructed to implement scanner electronics. The next scanner version with anticipated inverse scattering resolution of 1 mm or less and reflection tomography resolution of 0.5 mm or less should improve performance. Clinical images to date and performance will be discussed.

2:20

1pBB3. Image reconstruction using eigenfunctions of a scattering operator. Robert Waag (Dept. of Elec. and Comput. Eng. and Dept. of Imaging Sci., Univ. of Rochester, Rochester, NY 14627)

A multiple-frequency iterative inverse scattering method is reviewed. In the method, scattered acoustic fields are described using a scattering operator with eigenfunctions that correspond to far-field patterns of an effective source distribution. Incident-wave patterns determined using the eigenfunctions focus on the distribution. The focusing properties of the eigenfunctions are employed to reconstruct an object by using products of numerically calculated fields defined by weighted products of the eigenfunctions. The full range of frequencies present in an incident pulse waveform is employed. Iteration using a linearized version of the method and calculated scattering from an estimate of the object permits reconstruction of large-scale high-contrast objects. The method is illustrated using calculated and measured data. In the calculations, an exact solution for scattering from nonconcentric lossy cylinders with parallel axes was employed to obtain the scattered field. In the measurements, a novel ring-transducer system was used to obtain the incident and total fields. The results of simulations and experiments show that the method converges and is accurate for high-contrast large-scale tissue-mimicking objects. Reconstructions using measured scattering from random media have considerably less speckle than conventional b-scan images, implying that the inverse scattering method can clearly demonstrate small high-contrast inclusions.

3023   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3023**

1pBB4. Optical coherence-based imaging and sensing in biomedicine and biotechnology. Joseph A. Izatt (Duke Univ., Campus Box 90281, Durham NC 27708)

Optical coherence-based imaging techniques including optical coherence tomography (OCT), optical coherence microscopy (OCM), and spectral domain phase microscopy (SDPM) use low-coherence spectral interferometry to obtain nanometer to micron-scale measurements of structure, motion, and molecular composition in living cells, tissues, and organisms. OCT has become a standard diagnostic tool in clinical ophthalmology and is under investigation for other human diagnostic applications including cancer detection and evaluation of cardiovascular disease. Within the past few years, dramatic technology advances have increased the performance of OCT and OCM systems manyfold, and they are now capable of micron-scale two- and three-dimensional functional and molecular imaging noninvasively in living systems. Applications of these new technologies for high-throughput noninvasive phenotyping and rapid 3D imaging in small animals and developmental biology models is particularly compelling. Related technology advances have enabled the design of highly phase-stable interference microscopes capable of resolving nanometer-scale structures and motions in living cells with ms temporal resolution. These new capabilities are being used to probe cellular internal and external surfaces and their responses to chemical and mechanical stimuli. I will review our ongoing work at Duke in these areas. [Work supported by NIH Grants RR019769, EY013516, EB006338, and EY017393.]

1pBB5. Whole breast imaging with ultrasound tomography: First clinical results at the Karmanos cancer institute. Neb Duric, Peter Littrup, Earle Holsapple, and Olsi Rama (Karmanos Cancer Inst., 4100 John R, Detroit, MI 48201)

Although mammography is the gold standard for breast imaging, it leads to a high rate of biopsies of benign lesions and a significant false negative rate for women with dense breasts. A clinical breast-imaging device, based on the principles of ultrasound tomography, is being developed in response to this imaging performance gap. The computed ultrasound risk evaluation (CURE) system has been designed with the clinical goal of providing fast, whole-breast, operator-independent imaging, capable of differentiating breast masses. In an initial assessment of its *in vivo* performance, 50 patients were examined with CURE. The following results were demonstrated: (*i*) tomographic imaging of breast architecture, (*ii*) in-plane spatial resolution of 0.5 mm in reflection and 4 mm in transmission, (*iii*) imaging of breast masses in reflection and transmission mode (sound speed and attenuation, and (*iv*) differentiation of breast masses was shown to be feasible. These initial results indicate that operator-independent, whole-breast imaging and the detection and characterization of breast masses are feasible. Future studies will focus on improved detection and differentiation of masses with the long-term goal of improving the specificity of breast exams. [The authors acknowledge financial support from the Karmanos Cancer Institute and the NIH (grant R43 CA108072-01).]

1pBB6. Advanced tomographic algorithms and their potential applications to ultrasound imaging. Xiaochuan Pan, Samuel LeRoque, Emil Sidky (Dept. of Rad., MC2026, The Univ. of Chicago, 5841 S. Maryland Ave., Chicago, IL 60637, xpan@uchicago.edu), and Mark Anastasio (Illinois Inst. of Technol.)

The success of computed tomography has led to the development of a wide variety of tomographic imaging techniques. There has always been a great interest in ultrasound-based tomographic imaging because it can provide detailed, quantitative information about the imaged subject, which can be used for diagnosis, evaluation, and monitoring of cancer and other diseases. Tomographic imaging relies not only upon the imaging system for data acquisition but also upon the algorithms for image recovery from the data. In the last decade or so, there has been a tremendous development of innovative algorithms for tomographic imaging. One important breakthrough is that an artifact-free ROI image within the subject can be obtained from, e.g., truncated data. This breakthrough has led to the development of the so-called targeted-ROI imaging strategy in which one can substantially reduce imaging effort and artifacts. There have also been significant advances in reconstructing images from highly incomplete data acquired, e.g., at a small number of views and/or over a limited angular range. In this paper, we will discuss these new developments and their potential applications to ultrasound-based imaging. We will use examples to illustrate how they can be exploited for reconstructing images in different types of ultrasound-based tomographic imaging.

Contributed Papers

1pBB7. On the use of inverse scattering tomographic methods for quantitative ultrasound techniques. Roberto J. Lavarello and Michael L. Oelze (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

The application of the distorted Born iterative method (DBIM) is proposed to construct tomographic images of small scale structures. Specifically, DBIM is proposed as a tool to construct better ultrasonic models of cells in order to improve the accuracy of quantitative ultrasound (QUS) analysis. An alternative implementation of DBIM valid for measurement configurations where the receiver positions are fixed is discussed. In addition, a method for the selection of the regularization parameter is deter-

mined, based on the Rayleigh quotient iteration. QUS imaging has been used to differentiate malignant from benign tissue. Recently, cells have been hypothesized as an important source of scattering in tissue. Generating tomographic images of sound speed and density of cells would allow better models to be constructed. Because the contrast mechanism in cells is unknown, the effect of variable density is explored. Stage II–III Xenopus oocytes are proposed as an initial cell model for scattering, with typical sizes ranging between 300 and 600 μm. Based on measurements of the propagation delay through oocytes, it was determined that their diameters correspond to at most 24 resolution cells when using single-frequency DBIM. Hence, frequency hopping may be required to discriminate smaller features inside the oocytes.

4:10

1pBB8. Linear sampling method for acoustic inverse scattering in breast microcalcification detection.
Matthew A. Lewis, Peiying Liu, Edmond Richer, Peter P. Antich (Dept. of Radiol., UT Southwestern Medical Ctr. at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9058, matthew.lewis@utsouthwestern.edu), and Stuart G. Johnson (UT Austin, Austin, TX)

In women under age 40, the breast may be too dense for mammographic screening and microcalcification detection. A nonionizing method would be preferred for frequent screening in patients with significant risk factors for developing breast cancer. Ultrasonic detection of microcalcifications has been problematic since frequencies above 10 MHz suffer from appreciable attenuation in soft tissues. Transmission diffraction tomography is by construction a low-pass filter and insensitive to scattering caused by small, hard inhomogeneities. A more general form of acoustic inverse scattering is therefore needed for microcalcification detection and localization by ultrasound. An advanced scalar inverse scattering theory developed by Colton, Kirsch, and others [D. Colton, J. Coyle, and P. Monk. SIAM Review **42**(3), 369–414 (2000)] in the electromagnetic inverse scattering community for determining the location and shape of scatterers with size on the order of the wavelength (the so-called resonance region) has been simulated in the context of breast microcalcification detection. This method was shown to be related to both multiple signal classification and time-reversal algorithms [M. Cheney. Inverse Problems **17**, 581–585 (2001)] [S.K. Lehman and A.J. Devaney J. Acoust. Soc. Am. **113**(5), 2742–2753 (2003)]. Translation of this method to a heterogeneous soft-tissue background is a significant challenge.

4:25

1pBB9. Full wave nonlinear inverse scattering for transmission breast imaging.
James Wiskin (Techni Scan Medical Systems, 1011 Murray Holladay Rd., Salt Lake City, UT, 84117, jwiskin@techniscanmedical.com), David Borup, and Steven Johnson

Ultrasound has been proposed as an adjunct diagnostic modality for 25+ years. However, initial studies were hampered by relatively low resolution. Over the past 20 years there has been a tremendous amount of research carried out in inverse scattering, the mathematical foundation behind quantitative determination of the coefficients of a partial differential equation. In particular, the research at TechniScan, Inc., and the Dept. of Bioengineering, Univ. of Utah, has culminated in algorithms that give approximately millimeter resolution, and quantitative accuracy, in 2-D and 3-D problems. We invert the full nonlinear acoustic inverse scattering problem (no Born/Rytov approximation) in 2-D and 3-D. We reconstruct a phantom with known inclusions. The image space is $241 \times 241 \times 135 = \sim 7.8$ Megapixels or larger. Each pixel is $0.8 \times 0.8 \times 1.0 mm^3$ in size. The data is collected on a 6 by 160 element array, positioned 180 views surrounding the breast. The transmitted wave is a plane wave chirp. Reconstruction time is ~ 2 h on a six-node cluster of Pentium Ms. This paper will discuss theoretical algorithm development, practical issues that result in images, and a device presently being used in a clinical setting. We show several reconstructed images from ~ 400 volunteers.

4:40

1pBB10. Radial reflection diffraction tomography.
Sean K. Lehman (Lawrence Livermore Natl. Lab., L-154, 7000 East Ave., Livermore, CA 94550) and Stephen J. Norton (Geophex, Ltd., Raleigh, NC 27603)

A wave-based tomographic imaging algorithm based upon a single, rotating, radially outward oriented transducer has been developed. At successive angular locations at a fixed radius, the transducer launches a primary field and collects the backscattered field in a "pitch/catch" operation. The hardware configuration, operating mode, and data collection method are identical to that of most medical intravascular ultrasound (IVUS) systems. The goal is to develop a wave-based imaging algorithm using diffraction tomography techniques. Given the hardware configuration and the imaging method, this system is referred to as "radial reflection diffraction tomography." Two hardware configurations are considered: a multimonostatic mode using a single transducer as described

above, and a multistatic mode consisting of a single transmitter and an aperture formed by multiple receivers. The probe is mounted at the end of a catheter or snaking tube inserted into a part or medium with the goal of forming images of the plane perpendicular to the axis of rotation. An analytic expression for the multimonostatic inverse is derived but ultimately a Hilbert space inverse wave algorithm is used to construct images using both operating modes. Applications include improved IVUS imaging, bore hole tomography, and nondestructive evaluation (NDE) of parts with existing access holes.

4:55

1pBB11. Ultrasonic reflection imaging with a heterogeneous background for breast cancer detection and diagnosis.
Lianjie Huang (Los Alamos Natl. Lab., M.S. D443, Los Alamos, NM 87545), Youli Quan (Stanford Univ., Stanford, CA 94305), Nebojsa Duric, and Cuiping Li (Karmanos Cancer Inst., Detroit, MI 48201)

Sound speeds and densities of breast tissue vary in space, and those of tumors are different from the surrounding tissues. These differences in mechanical properties cause scattering of ultrasonic waves, particularly in dense breasts. An ultrasonic reflection imaging method using a heterogeneous background is presented to properly account for scattering off heterogeneities within the breast. The method is based on downward continuation of ultrasonic wave fields in the frequency-space and frequency-wave-number domains. Synthetic ultrasonic data for scatterers within a heterogeneous breast model were used to investigate the imaging resolution of the method. Imaging examples using ultrasonic data acquired using a clinical prototype, consisting of an ultrasound scanner with a ring transducer array, are presented. The heterogeneous background sound-speed models were obtained using transmission tomography. The imaging results are compared with those obtained using a homogeneous background. The study demonstrates that imaging with heterogeneous backgrounds significantly improves image quality and resolution. [This work was supported through the DOE Laboratory-Directed Research and Development program at Los Alamos National Laboratory.]

5:10

1pBB12. Photoacoustic imaging with a planar measurement array: Improving the resolution using image sources.
Benjamin T. Cox (Dept. of Med. Phys. and Bioengineering, Univ. College London, Gower St., London, WC1E 6BT, UK, bencox@mpb.ucl.ac.uk), Simon. R. Arridge, and Paul C. Beard (Univ. College London, Gower St., London, WC1E 6BT, UK)

Biomedical photoacoustic (PA) imaging is a soft-tissue imaging modality that combines the high spatial resolution of ultrasound (US) with the contrast and spectroscopic opportunities afforded by imaging optical absorption. Exact reconstruction algorithms for PA imaging have been described for spherical, cylindrical, and planar sensor arrays. Many applications to date have used a cylindrical measurement geometry because a single US transducer may be scanned straightforwardly in a circle around a sample. However, planar US arrays composed of piezo or optical elements with small element sizes and fast acquisition times are readily available, making them an attractive option for imaging applications. To capture sufficient data for an exact PA reconstruction with a planar geometry requires an infinitely wide array. In practice, of course, it will be finite, resulting in a loss of resolution and introducing artefacts into the image. To overcome this limitation, it is proposed that acoustic image sources are used to generate a periodically repeating sound field that can be reconstructed exactly from measurements made over a finite aperture. [This work was funded by the EPSRC, UK]

5:25

1pBB13. An acoustic technique to assess particle distribution following transcutaneous needle-free injection. Jamie Condliffe, Jamie R. T. Collin, George Costigan, and Constantin-C. Coussios (Inst. of Biomed.. Eng., Dept. of Eng. Sci., Univ. of Oxford, 17 Parks Rd., Oxford OX3 0DH, UK, constantin.coussios@eng.ox.ac.uk)

Needle-free injection is a novel technique enabling pain-free transdermal drug and vaccine delivery. The present work aims at developing a noninvasive, ultrasonic technique to assess distribution of drug particles delivered in this manner beneath an optically opaque layer, such as the skin. Drug particles are modeled as a monodisperse distribution of 50-μm polystyrene spheres that are injected into an agar-based tissue-mimicking material, which they penetrate to a mean depth of 500 μm. A 15-MHz broadband focused immersion transducer, driven in pulse-echo mode, was used to scan the surface of the agar phantom. By assuming that the mean-squared voltage of the received waveform is proportional to the number of particles encountered by the incident pulse, a 3-D map of particle distribution was generated. The resulting acoustic microscopy data were validated by examining phantom slices of thickness similar to the acoustic pulse length under a conventional optical microscope. The acoustic technique was found to provide a good prediction of the mean particle penetration depth and an accurate map of transverse particle distribution in planes normal to the direction of injection. It thus shows great promise as a noninvasive means of assessing the efficacy of needle-free injection, both *in vitro* and *in vivo*.

5:40

1pBB14. Speckle reduction of echographic image by transmitting ultrasonic waves of double peak spectrum. Iwaki Akiyama (Shonan Inst. of Technol., 1-1-25 Tsujido-nishikaigan, Fujisawa 251-8511 Japan), Shigemi Saito (Tokai Univ., Shimizu-ku, Shizuoka, 424-8610 Japan), and Akihisa Ohya (Univ. of Tsukuba, Tsukuba, Ibaraki, 305-8573 Japan)

Speckle noise of ultrasonic echography is a major issue. To reduce the speckle noise a method of summation after logarithmic compression of higher order harmonic images is proposed. The feasibility is confirmed by the phantom experiments using a mechanical scanner system equipped with a bi-layered structure transducer of PZT and PVDF. The dynamic range of harmonic components higher than the third, however, is not enough for effective speckle reduction. In this study a method of using difference and sum frequency components as well as harmonic components by transmitting pulsed waves of double peak spectrum is proposed. To compare the dynamic range of those components the numerical simulation of transmitting the ultrasonic waves of 2 and 6 MHz and those of 2 and 8 MHz in absorbing media is carried out. An attenuation coefficient of 1.0 dB/cm/MHz is assumed. As a result, complex beam patterns are formed due to degeneration between second harmonic and difference frequency components of 4 MHz in the former case. In the latter case three focal beam patterns of 4-, 6-, and 8-MHz components within 40 dB in dynamic range are formed. Consequently this method has the potential to apply for the biological tissues. [Work supported by a Grant-in-Aid for Scientific Research.]

TUESDAY AFTERNOON, 28 NOVEMBER 2006　　　　　　　OAHU ROOM, 1:00 TO 4:00 P.M.

Session 1pEA

Engineering Acoustics, Acoustical Oceanography and Underwater Acoustics: Acoustic Sub-Bottom Profiling Technology

Kenneth M. Walsh, Chair

K. M. Engineering Ltd., 51 Bayberry Ln., Middletown, RI 02842

Chair's Introduction—1:00

Invited Papers

1:05

1pEA1. Subbottom profiling using synthetic aperture processing. Steven Schock (Dept. of Ocean Eng., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431, schock@oe.fau.edu)

Using synthetic aperture processing, a chirp sonar generates reflection profiles of the seabed. Synthetic aperture processing is implemented by calculating the vertical change in the positions of acoustic transducers between transmission and using the offsets to time shift the acoustic data so that horizontal reflectors are summed coherently. The calculated changes in the positions of the acoustic transducers between transmission events are based on Doppler velocity log and inertial measurement unit measurements of vehicle motion. Imagery generated from reflection profiles of seabeds with various sediment types are used to quantity the improvement in the spatial resolution and the SNR of reflection profiles as a function of the length of the synthetic aperture. [Work supported by Office of Naval Research.]

1:25

1pEA2. Chirp seismic surveying of shallowly buried fluvial systems on the outer New Jersey continental shelf: Mapping channel morphology and internal stratigraphy in 3D. John Goff, Sylvia Nordfjord, James Austin, Jr. (Inst. for Geophys., Jackson School of Geosciences, Univ. of Texas, Austin, TX 78759), and Steven Schock (Florida Atlantic Univ.)

The outer New Jersey continental shelf has been the focus of Office of Naval Research-sponsored acoustic studies for more than a decade. As part of this effort, ONR has funded a number of field programs to characterize the geologic history and sedimentary properties of the shallow seabed. We report here on a chirp seismic campaign conducted in 2001, with particular focus on characterizing the morphology and internal stratigraphy of shallowly buried fluvial channel systems that are widespread in this region. These channels were formed during the last glacial lowstand, and then filled and buried by sediments during sea level rise. Mapping the channel morphology in 3D presented a significant challenge. Although coverage (200-m line spacing) was dense, application of

standard interpolation techniques resulted in poor representation of channel continuity along sinuous pathways. We developed a technique for transforming the seismic horizon data into a channel-oriented reference frame, wherein the channel is effectively straight and simple interpolation techniques can be used to maintain channel continuity. Seismic units within the channel fill could be discerned based on bounding reflectors and acoustic character and mapped throughout the 3D channel systems. These units provide a record of sea level change and drowned-river estuary development.

1:45

1pEA3. Removing multiples and depth imaging and inverting primaries without knowing or determining the velocity model: A response to pressing seismic challenges. Arthur B. Weglein (Phys. Dept., Univ. of Houston, 4800 Calhoun, Houston, TX 77004, aweglein@uh.edu)

The isolated task specific subseries of the inverse scattering series represents a direct response to many pressing seismic exploration challenges. Those challenges often involve removing multiply reflected events and locating and identifying targets when the subsurface properties and boundaries are rapidly varying and beyond all current methods to adequately predict and estimate. The goal of these fundamentally new concepts and algorithms is to make currently inaccessible targets accessible and accessible targets better defined. We will exemplify these new algorithms with synthetic and field data. Open issues and plans will be discussed.

2:05

1pEA4. Seismic imaging of subduction zone deformational structures using 3D seismic profiling. Nathan L. Bangs (Inst. for Geophys., Univ. of Texas, 4412 Spicewood Springs Rd., Austin, TX 78759, nathan@ig.utexas.edu)

In the past 15–20 years, considerable efforts have been made to image complex deformational structures beneath the seafloor at subduction zones using seismic imaging techniques. The goals have been to: 1) accurately image structures to determine the geometry of thrust faults controlling the deformation and tectonics of plate collision at subduction zones, and 2) to determine the rock physical properties from seismic reflections to infer fault zone mechanical properties and examine conditions leading to fault rupture and earthquake generation. Subduction zones generate the world's largest and most destructive earthquakes and tsunamis and accurate seismic images of faults to depths of 6–7 km below seafloor offer an opportunity to examine conditions at these inaccessible depths. Three-dimensional seismic techniques have proven to be critical for sampling the reflected wavefield at the sea surface sufficiently to be able to migrate reflections to their proper subsurface positions and construct accurate images. I will present examples of these 3-D seismic reflection images, analyses of these data, and future directions. [Work supported by NSF.]

2:25

1pEA5. Application of CHIRP data to neotectonic, sediment dispersal and sea level studies. Elizabeth Johnstone and Neal Driscoll (Scripps Inst. of Oceanogr., Univ. of California San Diego, La Jolla, CA 92093-0244)

CHIRP seismic data will be presented in order to demonstrate the importance of subbottom imaging in addressing fundamental problems in Earth sciences. Examples from actively deforming regions in California place important constraints on the rupture history along major fault systems as well as tectonic control on sediment preservation and exposure of hardgrounds on the seafloor. Tectonic deformation creates a rich tapestry of geologic substrate exposed on the seafloor, ranging from mobile sands to lithified hardgrounds. It is this heterogeneity of the seafloor that creates different biological ecosytems and consequent increased biodiversity. CHIRP data have also provided new insights into sediment dispersal and clinoform development in the Gulf of Papua. In the northern gulf, oblique clinoforms suggest the rate of sediment supply has outstripped the formation of accommodation, whereas in the south the sigmoidal package indicates that accommodation increased faster than sediment supply. The morphology and internal geometry of the clinoform itself are inherently three-dimensional. The underlying physiography also appears to play an important role in governing sediment dispersal and distribution.

2:45−3:00   Break

Contributed Papers

3:00

1pEA6. A multibeam subbottom profiler. Jerald W. Caruthers (Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Center, MS 39529) and Sean Griffin (Omni Technologies, Inc., Stennis Space Center, MS 39529)

A multibeam subbottom profiling system (MSBPS) is being built with funding from the Defense University Research Instrumentation Program (DURIP) in coordination with the Office of Naval Research. The MSBPS is being constructed from three T70 Neptune Sonar transducer modules in a line. Each T70 has five 11.5-kHz elements in each of two parallel lines. The resulting system is composed of 15 elements in each line with a total length of 1.3 m separated by about 8.5 cm. Each line has full beamwidths to the 3-dB points of about 5×70 deg. The system will be used in an interferometric mode between the two lines. [This work is supported by ONR 321OA.]

3:15

1pEA7. Using ambient noise and vector sensor arrays to form a subbottom profiler. Paul Hursky and Martin Siderius (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130)

Surface-generated ambient noise is distributed in time, but its correlation produces a strong impulse that is excellent for imaging the ocean floor. This technique has been experimentally demonstrated using vertical line arrays of pressure phones. The up-looking endfire beam (for ambient noise traveling down) is cross-correlated with the down-looking endfire beam (for the ambient noise reflected from the bottom, traveling up). Each layer in the bottom produces a distinct reflection. Previous work has required somewhat large vertical arrays. A towed horizontal array would be much better and allow us to make measurements with a mobile platform. The problem is that a horizontal line array cannot distinguish up from down—a broadside beam is symmetric with respect to the axis of the array. Vector sensors, sensing both pressure and particle velocity, have additional degrees of freedom for overcoming such ambiguities. We will show how vector sensors make it possible to implement such an ambient

noise subbottom profiler, either in a greatly reduced vertical array or in a horizontal (towed) array configuration. We will also show results of processing data collected on a four-element vector sensor array and a much larger pressures-only array during the Makai experiment.

baffles and waveguides was also investigated. A theoretical discussion is also included. This work represents the practical employment of MHD transducers to undersea digital communication. [Work supported by General Atomics.]

3:30

1pEA8. Development of practical broadband magnetohydrodynamic transducers for undersea digital communication. Eric Berkenpas and Corey Jaskolski (Hydro Technologies, 700 Automation Dr., CO 80550)

Demand for broadband underwater digital communication is being driven by development of tetherless autonomous underwater vehicles (AUVs). Current communications methods employ typically piezoelectric transducers. While efficient, and broadband as receivers, these transducers exhibit significant frequency dependence when transmitting, thus limiting their utility in digital communications. Magnetohydrodynamic (MHD) transducers generate acoustic waves in seawater by applying time-varying Lorentz force to moving ionic charges in the presence of a magnetic field. MHD transducers are ultracompliant, exhibit wide-frequency bandwidth, and are therefore attractive for high-speed undersea digital communications. This paper reports on exploratory development MHD transducers with the goal of improved transduction efficiency and bandwidth. Novel MHD transducer structures were fabricated with permanent magnet, parallel electrode structures, and planar interdigitated electrode structures. To evaluate transducer performance, an automated transducer test setup was assembled incorporating a 7500-L saltwater test environment, a time-gated high-current signal source, and a wideband hydrophone. Use of acoustic

3:45

1pEA9. Influence of nonlinear propagation on calibration of hydrophone sensitivity using a two-transducer reciprocity method. Masahiro Yoshioka and Sojun Sato (Natl. Metrology Inst. of Japan, Natl. Inst. of Adv. Industrial Sci. and Technol., Tsukuba Central 3, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8563 Japan)

This paper describes the phenomenon by which the measured value depends on the ultrasound pressure in the measurement field when the hydrophone sensitivity has been calibrated using the two-transducer reciprocity method. This dependence is attributable to nonlinear ultrasonic propagation. However, the international standard, IEC 60866:1987, "Characteristics and calibration of hydrophones for operation in the frequency range 0.5 MHz to 15 MHz," provides a correction factor for diffraction and attenuation derived assuming linear propagation of sound. We attempted to derive the correction factor from the ultrasound pressure distribution calculated using the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation, which is well known as a nonlinear ultrasonic model. Results demonstrated that the dependence of the calibrated sensitivity value on measured sound pressure is quantitatively explainable. The hydrophone sensitivity can be calibrated properly by compensating for the influence of ultrasonic nonlinear propagation using the KZK equation.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

IAO NEEDLE/AKAKA FALLS ROOM, 1:30 TO 4:15 P.M.

Session 1pMU

Musical Acoustics: Singing Voices

Peter L. Hoekje, Cochair
Baldwin-Wallace College, Dept. of Physics and Astronomy, 275 Eastland Rd., Berea, OH 44017

Naotoshi Osaka, Cochair
Tokyo Denki Univ., Dept. of Information Systems and Multimedia Design, School of Engineering, 2-2 Kandanishiki-cho, Chiyoda-ku, Tokyo, 101-8457, Japan

Contributed Papers

1:30

1pMU1. Voice quality of artistic expression in Noh: An analysis-synthesis study on source-related parameters. Hideki Kawahara (Faculty of Systems Eng., Wakayama Univ., 930 Sakaedani, Wakayama, 640-8510 Japan), Osamu Fujimura (Intl. Inst. for Adv. Studies, Sorakugun, Kyoto 619-0225 Japan), and Yasuyuki Konpaku (Konparu School of Noh, Nara, Japan)

High-quality recordings of Noh vocalizations, produced by one of the authors, were made in an anechoic chamber, comparing vocal styles and emotional expression, using the same sample text from classical Noh. The recorded signals were subject to analysis, synthesis, and morphing by using STRAIGHT. The morphing procedure generated mean parameter settings for comparing two subtly differentiated voicing styles, traditionally called "wagin" soft vocalization, and "gougin," strong vocalization, to investigate the relation between signal parameters and communicative functions. STRAIGHT extracts three parameters from recorded speech signals: fundamental frequency, smoothed spectrogram, and a time-frequency aperiodicity map. The fundamental frequency and the time-frequency aperiodicity, in other words, source-related parameters, were systematically manipulated to investigate their role in artistic expression.

1:45

1pMU2. Spectro-temporal analysis of Palauan chants. Birgit Abels (Ruhr Univ., Imam Husnu Sok. 1/1, IStanbul Bochum, Germany) and Jonas Braasch (Rensselaer Polytechnic Inst., NY)

A field study was carried out to determine how songs from the repertory of chants in Palau have changed over the last century. To this end, several chanters were recorded during recent fieldwork; these sound documents were then compared to items from several historic collections of sound recordings from the Phonogram Archives Berlin, Germany, and the Pacific Collection, Library of the University of Hawai'i at Manoa. The available recordings cover the time span from 1909–2006. Analytical results were derived in the form of spectrograms both for recent recordings from the years 2005/2006 and historic ones from 1963. They confirmed the auditory impression that chanting in Palau today has moved away from a recitative style, relying predominantly on the speaking voice, towards a performance practice the sound ideal of which is the singing voice. Also, the range of pitch employed in these chants has expanded. Spectral analysis also revealed how, in particular genres, the musical form of group chants is structured by the changing contours of a frequency band that is

resolved at phrasal ends, when the singers narrow the range of this tonal cluster. This unique technique is one of the core musical elements of the Palauan singing tradition.

2:00

1pMU3. Spectral modification technique in conversion of musical notes or tempos for singing voice synthesis system. Hideki Banno, Kumi Ohta, Masato Kawazoe, and Fumitada Itakura (Faculty of Sci. and Technol., Meijo Univ., 1-501, Shiogamaguchi, Tenpaku-ku, Nagoya-shi 468-8502, Japan, banno@ccmfs.meijo-u.ac.jp)

The STRAIGHT time-frequency representations (spectrograms) of singing voice signals in various musical notes and various tempos are observed to develop a high-quality synthesis system of singing voice. The spectrogram of STRAIGHT, which is a very high-quality analysis-synthesis system, can represent the vocal tract information accurately. A conversion system of a musical note or a tempo of an input singing voice signal has been implemented based on the observation. As a result of the observation, the frequency warping of the STRAIGHT spectrogram based on a DP matching algorithm has been introduced into the system. It was found that the method using a differential of a smoothed spectrum as a spectral distance measure in the frequency warping produces subjectively better quality than that using a smoothed spectrum directly. It was also found that the method without spectral modification, i.e., only with pitch/tempo modification in the conversion, produces better quality than that using a differential of a smoothed spectrum. This can be caused by the destruction of naturalness in the method using a differential of a smoothed spectrum.

2:15

1pMU4. Physical modeling vocal synthesis. David Howard, M. Jack Mullen, and Damian Murphy (Audio Lab., Intelligent Systems Res. Group, Dept. of Electron., Univ. of York, Heslington, York YO10 5DD, UK. dh@ohm.york.ac.uk)

Physical modeling music synthesis produces results that are often described as being "organic" or "warm." A two-dimensional waveguide mesh has been developed for vocal synthesis that models the adult male oral tract. The input is either the LF glottal source model or a user-provided waveform file. The mesh shape, which is based on MRI data of a human vocal tract, can be changed dynamically using an impedance well approach to allow sounds such as diphthongs to be synthesized without clicking. The impedance well approach enables the mesh shape to be varied without removing or adding elements, actions that cause audible clicking in the acoustic output. The system has been implemented as a real-time MIDI-controlled synthesizer, taking its inspiration from the von Kempelin speaking machine, and this will be demonstrated live as part of the presentation. The system is set up to allow a continuous glottal source to be applied that includes vibrato, and thus the real-time output it currently produces is close to being vocalized. It should be noted, though, that appropriate variation of the lip opening does produce a voiced bilabial plosive, demonstrating the potential for moving towards a full speech synthesis system in the future.

2:30

1pMU5. Analysis of acoustic features affecting singing-voice perception and its application to singing-voice synthesis from speaking-voice using STRAIGHT. Takeshi Saitou (ATR Cognit. Information Sci. Lab, 2-2-2 Hikaridai Seika-cho Souraku-gun, Kyoto 619-0288, Japan), Masashi Unoki, and Masato Akagi (School of Information Sci. Japan Adv. Inst. of Sci. and Technol., Nomi Ishikawa 923-1292 Japan)

A singing-voice synthesis method that can be transformed from a speaking voice into a singing voice using STRAIGHT is proposed. This method comprises three sections: the F0 control model, spectral sequence control model, and duration control model. These models were constructed by analyzing characteristics of each acoustical feature that affects singing-voice perception through psychoacoustic experiments. The F0 control

model generates a singing-voice F0 contour through consideration of four F0 fluctuations: overshoot, vibrato, preparation, and fine (unsteady) fluctuation that affect the naturalness of a singing voice. The spectral sequence control model modifies the speaking-voice spectral shape into a singing-voice spectral shape by controlling a singer's formant, which is a remarkable peak of a spectral envelope at around 3 kHz, and amplitude modulation of formants synchronized with vibrato. The duration control model stretches the speaking-voice phoneme duration into a singing-voice phoneme duration based on note duration. Results show that the proposed method can synthesize a natural singing voice, whose sound quality resembles that of an actual singing voice.

2:45–3:00   Break

3:00

1pMU6. Automatic detection of head voice in sung musical signals via machine learning classification of time-varying partial intensities. Ryan J. Cassidy and Gautham J. Mysore (Dept. of Music, Ctr. for Comput. Res. in Music and Acoust., 541 Lasuen Mall, Stanford, CA 94305)

The automatic detection of portions of a musical signal produced according to time-varying performance parameters is an important problem in musical signal processing. The present work attempts such a task: the algorithms presented seek to determine from a sung input signal which portions of the signal are sung using the head voice, also known as falsetto in the case of a male singer. In the authors' prior work [Mysore *et al.*, Asilomar Conf. Signal. Sys. Comp. (2006) (submitted)], a machine learning technique known as a support vector classifier [Boyd and Vandenberghe, 2004] was used to identify falsetto portions of a sung signal using the mel-frequency cepstral coefficients (MFCCs) of that signal (computed at a frame rate of 50 Hz). In the present work, the time-varying amplitudes of the first four harmonics, relative to the intensity of the fundamental, and as estimated by the quadratically interpolated fast Fourier transform (QIFFT) [Abe and Smith, ICASSP 2005], are used as a basis for classification. Preliminary experiments show a successful classification rate of over 95% for the QIFFT-based technique, compared to approximately 90% success with the prior MFCC-based approach. [Ryan J. Cassidy supported by the Natural Sciences and Engineering Research Council of Canada.]

3:15

1pMU7. A proposal of discrimination method between voice and music using Gaussian mixture models and similarity in a music selection. Takuma Yoshida, Masaaki Ishigame, Yoshiaki Ito, and Kazunori Kojima (Iwate Prefectural Univ., Sugo 152-52, Takizawa, Iwate, 020-0193, Japan, g231e038@edu.soft.iwate-pu.ac.jp)

We have proposed a method for automatic music boundary detection using similarity in a music selection. The method is able to capture a whole position of a consecutive music selection, although the boundary position of a music selection is likely to be roughly estimated. In this paper, we propose a new method that combines a GMM (Gaussian mixture model) discrimination method between music and voice with the previously proposed method. The GMM enables us to determine strict boundary positions. On the other hand, the GMM often misdetects boundaries that are not actual boundaries, such as changing points of music instruments, the point of modulation in a music selection, and so on. We, therefore, exclude the GMM misdiscrimination in a music selection by the previously proposed method, and also realize the precise detection of music boundaries by the GMM. We conducted various experiments using open music selections that are provided by RWC, and the results showed the proposed method could improve the performance of music boundary detection, compared with the previous method and the method using only the GMM.

3:30

1pMU8. Acoustical analysis of nonlinguistic sounds. Yell Inverso, Corine Bickley, and Charles Limb (Gadlaudet Univ., 800 Florida Ave. NE, Washington, DC 20011, danielle.inverso@gallaudet.edu)

The perception of speech is an undisputed important goal for cochlear implantation; however, the reception of nonlinguistic sounds (NLS) is also important. Nonlinguistic sounds are important for one's safety and they are considered to have importance to a person's sense of connection to and welfare in one's environment. NLS are different from speech, and the perception and acoustic characteristics of NLS have not been studied adequately to allow clinicians to fit CIs for optimal recognition of both speech and NLS. The specific categories of nonlinguistic sounds to be evaluated in this study are (1) human vocal/nonverbal, such as a baby crying or a person coughing; (2) mechanical/alerting, such as a telephone or an alarm clock buzzing; (3) nature/inanimate, such as weather sounds; (4) animal/insect, such as a dog barking; and (5) musical instruments such as the strum of a guitar. In this ongoing study, 50 listeners (half adults with normal hearing and half adult post-ling deafened CI users) are being evaluated with respect to their ability to identify the sources of a variety of common, everyday NLS. In this poster, an acoustic analysis of a representative set of the stimuli will be presented.

3:45

1pMU9. A psycho-acoustical analysis of vocal effort by opera singers. Konrad Skirlis, Densil Cabrera, and Anna Connolly (P.O. Box A2106, Sydney South 1235 NSW, Australia, audiodarnok@yahoo.com)

A singer's vocal output will be affected by a number of factors such as the size of the auditorium used for the performance. In this paper, the "large hall" and "small hall" voices of singers were explored in terms of the specific loudness pattern, which represents the distribution of loudness

in the auditory system. Recordings of seven professional opera singers were made in an anechoic room and in a small recital hall. Singers were asked to imagine they were in a small hall for one set of recordings and in a large hall for another set. Each singer sang the same song under these two "conditions." The recordings were then analyzed to determine spectral and temporal changes. Spectral analysis was used to calculate the strength of the singers formant under these conditions, and this was compared with the specific loudness patterns. Results help us to understand why the singer's formant is so important for vocal projection.

4:00

1pMU10. The effect of seating arrangement on choral sound in a nonselected mixed collegiate choral ensemble. Wei-Chun Wang (St. Johns Univ., 499, Sec. 4, Tam King Rd., Tamsui, Taipei, 25135 Taiwan, R. O. C.)

With both macroarrangement concerns and microarrangement concerns, choral directors use singer's placement as an essential technique to find compatible voices by listening to individual voices in various combinations. This study investigated the effect of seating arrangements, musical textures, and language familiarity on the choral balance and choral blend in an inexperienced collegiate choral ensemble. After ten rehearsals directed by the investigator, 48 nonselected choristers of a curricular SATB choral ensemble sang four musical excerpts, with the combination of two musical textures (homophonic and polyphonic compositions) and two types of languages (native and foreign language), a cappella in three formations (a sectional block, a mixed quartet, and a hybrid formation) in a chapel. Five experienced judges evaluated the screened live performance by rating the choral balance and choral blend on seven-point scales. The process was repeated. For blend only, sectional and hybrid formations were favored for polyphonic excerpts. The excerpts sung in native language were rated higher than the foreign excerpts, in particular for blend.

TUESDAY AFTERNOON, 28 NOVEMBER 2006

NIIHAU ROOM, 1:00 TO 4:40 P.M.

Session 1pPAa

Physical Acoustics and Engineering Acoustics: Infrasound

Milton A. Garces, Chair

Univ. of Hawaii, Infrasound Lab., 73-4460 Queen Kaahumanu Hwy., Kailua-Kona, HI 96740-2638

Chair's Introduction—1:00

Invited Papers

1:05

1pPAa1. Contribution of continuous infrasound monitoring to scientific research. Alexis Le Pichon, Karl Antier (CEA/DASE BP12, Bruyeres-le-Chatel, France), Lars Ceranna (Federal Inst. of Geosciences and Natural Resources, 30655 Hannover, Germany), Doug Drob (NRL, Washington, DC 20375), and Michel Lardy (IRD Ctr., 98848 Noumea Cedex, New-Caledonia)

Continuous automatic processing of the data is being performed at the French NDC from a number of IMS-type infrasound stations in order to detect and characterize coherent infrasonic waves recorded on the arrays. A large number of detections are associated with natural phenomena. Infrasonic waves generated by strong earthquakes occurring within region of high mountains are of great interest. Such events allow the evaluation of the relative contributions of the different source mechanisms involved in large earthquakes, and offer an opportunity to improve our understanding of the amplification of ground displacement caused by the topography. Microbaroms as well as mountain waves are quasipermanent sources of infrasonic waves detected on a global scale. Their monitoring reveals clear periodic trends in the detected bearings and signal amplitude, providing further confirmation that long-range propagation strongly depends on the atmospheric conditions. Signals from volcanic activities also offer a unique opportunity for atmospheric studies. At large propagation ranges from the volcanoes, infrasound measurements can be used as input of inversion procedures to evaluate more precisely the vertical structure of the wind in a range of altitude unaccessible to ground-based or satellite measurements.

1pPAa2. Infrasound associated with stratospheric ash injection, window-shattering explosions, and incandescent avalanches at Tungurahua Volcano, Ecuador. Milton Garces, David Fee, Sara McNamara (Infrasound Lab., Univ. of Hawaii, Manoa, 73-4460 Queen Kaahumanu Hwy., #119, Kailua-Kona, HI 96740-2638, Milton@isla.hawaii.edu), David McCormack (Geological Survey of Canada, Ottawa, ON K1A 0Y3, Canada), and Henry E. Bass (NCPA, University, MS 39688)

On July 14, 2006, Tungurahua volcano in Ecuador initiated a dramatic eruptive sequence that caused the evacuation of villagers, the destruction of crops, and the reports of ash clouds above 15 km. Acoustic signals have been recorded by the prototype infrasound acoustic-surveillance system (ASHE) currently operating in Ecuador. The infrasonic signals recorded by the ASHE included explosions, pyroclastic flows, and a sustained low frequency vibration of the ground. We present the features of the different types of eruption signals, and discuss the physics of eruptive processes and volcanic ash injection into the atmosphere. We also compare the July 2006 ash-rich eruption sequence with the May 2006 eruptive episode, which injected very little ash into the atmosphere. Results are of significance for the future operational use of acoustic monitoring of volcanic activity for aviation safety.

1:45

1pPAa3. Infrasound from tornadoes and other severe storms. John M. Noble (U.S. Army Res. Lab., 2800 Powder Mill Rd., Attn: AMSRD-ARL-CI-ES, Adelphi, MD 20783, jnoble@arl.army.mil) and Alfred J. Bedard, Jr. (NOAA/Earth System Res. Lab./PSD/CIRES, Boulder, CO 80305)

Infrasound arrays have been used for a long time to detect and monitor man-made and natural acoustic sources below 20 Hz. Much of the earlier work on infrasound from severe weather was below 0.1 Hz and was related to severe weather penetrating the tropopause. It was noticed when violent thunderstorms passed near an array that a coherent signal would be detected coming from the same direction as the storm. By correlating the measurements with visual, satellite, and Doppler radar observations, the characteristics of the storm could be studied to determine what type of storm would cause the production of infrasound. Most of the storms that would produce an infrasonic signal were the strong supercell-type storms that produced intense gust fronts and tornadoes. More recent measurements at frequencies above 0.1 Hz have been associated with tornadoes and tornadic storms. Following on these observations, NOAA established an infrasonic network to evaluate the use of infrasound for tornado detection and warning. This network has operated since 2003 and is still being optimized and evaluated. This talk will cover the studies that have been conducted and numerical models developed to try and explain this phenomenon.

2:05

1pPAa4. High-altitude infrasound propagation experiment. Bill Andre (U.S. Army Space and Missile Defense Command, P.O. Box 1500, Huntsville, AL 35807, andreb@smdc.army.mil) and Henry E. Bass (Natl. Ctr. for Physical Acoust., University, MS 38677)

Experimental verification of propagation predictions and sensor performance for infrasound frequencies that propagate long distance (greater than 100 km) are difficult to perform due to the rarity of controlled sources. If the source is elevated, the lower atmospheric pressure allows an explosion of a given charge weight to expand further during the supersonic phase, resulting in a lower fundamental frequency. A 50-lb charge exploded at 40 km yields a fundamental frequency similar to a 500-ton explosion at ground level. During the fall of 2005 and the winter and summer of 2006, experiments were conducted at White Sands Missile Range, NM, where an Orion rocket lifted a 50-lb explosive charge to altitudes between 30 and 40 km. Launches occurred near 0200 and 0600 MST. Portable infrasound arrays to record events were deployed in New Mexico, Texas, Arizona, and California to complement permanent arrays. Infrasound recordings were supplemented with extensive meteorological measurements. Data analysis is in a preliminary stage. The High Altitude Infrasound Propagation Experiment resulted from collaboration between several U.S. Army and Navy commands, University of Mississippi, Southern Methodist University, University of Hawaii, University of California at San Diego, University of Alaska at Fairbanks, BBN Technologies, SAIC, and Los Alamos National Laboratory.

Contributed Papers

2:25

1pPAa5. Modeling the propagation of volcanic infrasound with a 3-D parabolic equation. Joseph F. Lingevitch, Geoffrey F. Edelmann, Douglas P. Drob, and Michael D. Collins (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Active volcanoes are significant sources of infrasound in the 1–4-Hz frequency band and are detectable at ranges of hundreds of kilometers. The propagation is affected by the atmospheric temperature, wind, and attenuation profiles, which are range and time dependent. In this paper, a 3-D atmospheric parabolic equation method that accurately incorporates the effects of wind is applied to modeling volcanic infrasound propagation. A parabolic equation is derived using a narrow angle operator approximation in azimuth and a wide angle approximation in depth to account for wind-induced refraction that can trap or disperse sound energy, leading to large variations of propagation with wind conditions. The atmospheric specification for this model is derived from the attenuation coefficients of Sutherland and Bass [J. Acoust. Soc. Am. **115**, 1012–1032 (2004)] and the temperature/wind profiles from the Naval Research Laboratory Ground to Space model [J. Geophys. Res. **108**, 4680–4691 (2003)], which provides global atmospheric profiles from the ground to 200 km

altitude updated four times daily. Signal amplitude and direction of arrival estimates from the model are compared with measurements of volcanic infrasound at listening stations several hundred kilometers from the source. [Work supported by ONR.]

2:40

1pPAa6. Infrasonic scattering studies using time-domain parabolic equation (TDPE) and gravity wave models. David E. Norris (BBN Technologies, 1300 N. 17th St, Ste. 400, Arlington, VA 22209)

The time-domain parabolic equation (TDPE) model is useful in predicting infrasonic waveforms, as it can account for diffraction and scattering mechanisms. In this study, a split-step Fourier (SSF) implementation of the TDPE is used to study infrasonic propagation. Scattering is modeled using a spectral gravity wave model, and multiple gravity wave realizations are used to quantify the uncertainty in waveform properties introduced by scattering from the gravity wave inhomogeneities. TDPE predictions are compared to specific infrasonic events and conclusions regarding the dominant propagation mechanisms are presented.

2:55

1pPAa7. The deep sound of one wave plunging. Milton Garces, David Fee, Sara McNamara (Infrasound Lab., Univ. of Hawaii, Manoa, 73-4460 Queen Kaahumanu Hwy., #119, Kailua-Kona, HI 96740-2638), Jerome Aucan, and Mark Merrifield (Univ. of Hawaii, Manoa, Honolulu, HI 96822)

The ability to provide infrasound estimates of breaking ocean wave height and period in shallow reefs, steep rocky coastlines, and sand beaches has been demonstrated in previous work. Yet defining the source process and isolating the source pressure function remained elusive because of ambiguity introduced by complex coastlines and multiple breaker zones. Due to the steep bathymetry and the its proximity to land, the Temae reef in the northeast coast of Moorea island, French Polynesia, provided a well constrained experimental environment where individual breaking waves could be identified and recorded. Synchronous wave height, infrasonic, seismic, and visual recordings of individual waves breaking against the shallow reef ledge were made and correlated. We characterize a possible source mechanism for surf infrasound and demonstrate the capability to acoustically track alongshore traveling (peeling) plunging waves. We also introduce preliminary results on remote infrasonic monitoring of the surf zone in the North Shore of Hawaii during the 2006-07 Winter high surf season.

3:10–3:25 Break

3:25

1pPAa8. Low-frequency acoustic propagation in a shallow water internal wave field. Geoffrey F. Edelmann, Joseph F. Lingevitch, and David C. Calvo (U.S. Naval Res. Lab., 4555 Overlook Ave. SW, Code 7145, Washington, DC 20375)

A fully 3-D parabolic equation is used to simulate low-frequency acoustic propagation through a nonlinear internal wave field. Previous simulations have shown that acoustic propagation is dependent on the relative orientation between the direction of the internal wave train and the acoustic signal. When the acoustic and internal waves are parallel, a simple 2-D vertical refraction is the dominant interaction. As the two waves become perpendicular, horizontal refraction of the acoustic signal due to azimuthal coupling has been suggested [Frank *et al.*, J. Acoust. Soc. Am. **118**, 723–734 (2005)]. Comparisons of 3-D acoustic propagation through an internal wave field with and without azimuthal coupling will be made. The parabolic equation used makes a narrow angle operator approximation in azimuth and a wide angle approximation in depth. We will extend the previous work such that coupling in both the horizontal and vertical direction will be calculated simultaneously. [This work is supported by the Office of Naval Research.]

3:40

1pPAa9. Infrasonic wave observations at Fairbanks, AK, of Augustine volcano eruptions during January 2006. Charles Wilson, Steve McNutt, John Olson, and Guy Tytgat (Geophysical Inst., Univ. of Alaska, Fairbanks, AK)

The recent Augustine eruptions, from the 11th to the 28th of January 2006, have produced a series of ten infrasonic signals observed at the I53US array. The eruption times for the signals were provided by the Alaska Volcanic Observatory at UAF using a Chaparral microphone present on Augustine Island a few kilometers from the crater. The bearing and distance of Augustine from I53US are respectively 207.8 deg and 675 km. The analysis of the signals is done with a least-squares detector/estimator that calculates, among other things, the horizontal trace-velocity and the azimuth of arrival of the signal. The average values of the trace-velocity and azimuth for all ten Augustine signals are 0.336 ± 0.0136 km/s and 208.7 ± 1.5 deg. respectively. The celerity for each signal was calculated using the range 675 km and the individual travel times to I53US. The average celerity for all ten eruption signals was 0.287 ± 0.018 km/s. Ray tracing studies, using wind speed and temperature profiles supplied by Dr. Doug Drob at NRL, have shown that both stratospheric and thermospheric ray paths are present in the data set.

3:55

1pPAa10. Infrasound signals detected by the observation system in Isumi, Japan. Takahiko Murayama, Motohiro Honma, Mami Nogami, Takayuki Otsu, Takafumi Amano, Yasuharu Saitoh, Nobuo Arai, and Toyomi Sakamoto (Res. Dept., Japan Weather Assoc., 3-1-1 Higashi-Ikebukuro, Toshima-ku, Tokyo 170-6055 Japan)

An infrasound observation system is installed in Isumi, Japan (35.3N, 140.3E) as the component of the International Monitoring System for CTBT. The observation system was deployed on November 2004. It comprises a six-element array. Five elements of the array are deployed as a pentagon with sides approximately 1.2 km long. The sixth element is located within the pentagon-shaped array. Until now, some interesting infrasound signals were observed. One of them was the infrasound wave produced by the large earthquake that occurred off the Pacific coast of North-East Japan on 16 August 2005. The infrasound wave-train, which had a spindle-shaped envelope, was detected following the arrival of the seismic wave. The arrival direction of the infrasound wave, which had maximum amplitude, indicated the direction of the land area, which had maximum intensity. The time difference between the on-set time of that infrasound wave and the origin-time of the earthquake coincided with the travel time for propagation in the atmosphere between the area that had maximum intensity and the observation site. Therefore, it was estimated that the observed infrasound waves were produced by shaking ground. In this paper, some remarkable infrasound signals including the above are discussed.

4:10

1pPAa11. Infrasound wind noise contributions. Richard Raspet, Jiao Yu, and Jeremy Webster (Univ. of Mississippi, University, MS 38677)

In a recent study [J. Acoust. Soc. Am. **119**(2), 834–843 (2006)] predictions of various noise contributions were compared to measured wind noise from various devices and windscreens. The fluctuating stagnation pressure, the turbulence-turbulence interaction pressure, and the self-noise contributions in the inertial range were calculated from the measured wind velocity spectrum and compared to measurements. In that study the bare microphone agreed well with the predicted stagnation pressure spectrum, and a heavily windscreened microphone agreed well with the predicted turbulence-turbulence interaction spectrum. In this paper we extend the analysis and the measurements into the source region of the spectra—well into the infrasound range. In addition we predict the spectral contribution due to the turbulence-mean shear interaction. The prediction displays a steep slope in the inertial range and may dominate at low frequencies. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

4:25

1pPAa12. Comparitive study of wind noise reduction devices. Jeremy Webster, Richard Raspet, and Jiao Yu (Univ. of Missississppi, University, MS 38677)

Spherical porous foam windscreens are commonly used to reduce wind noise in acoustic measurements. We have argued [J. Acoust. Soc. Am. **119**(2), 834–843 (2006)] that the wind noise measured outdoors will lie between the predicted fluctuating stagnation pressure generated by the microphone interaction with the incoming air flow for bare microphones and the turbulence-turbulence interaction pressure for a very large wind screen. The pressure level measured in a moderate size screen will lie between these two extremes. Different transducer geometries and arrangements have been used in aerodynamic and meteorological studies to measure the "fluctuating static pressure" without interference from transducer generated wind noise. In this paper we compare measurements of wind noise using bare microphones, microphones in foam windscreens, streamlined probes, the Quad Disk, and microphones mounted in flat plates. The purpose of this study is to clarify the principles involved in wind noise generation by transducer-turbulence interaction. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

Session 1pPAb

Physical Acoustics: Sound Speeds, Phonons, and Thermodynamics of Condensed Matter

Keiji Sakai, Cochair
Univ. of Tokyo, Inst. of Industrial Science, 4-6-1 Komaba, Meguro-Ku, Tokyo, 153-8505, Japan

Cristian Pantea, Cochair
Los Alamos National Lab., MST-NHMFL, MS E536, Los Alamos, NM 87545

Chair's Introduction—1:00

Invited Papers

1:05

1pPAb1. Determination of elastic properties and interatomic force constants using inelastic neutron scattering. Robert McQueeney (Dept. of Phys. and Astron., Ames Lab., Iowa State Univ., Ames, IA 50011)

Neutron scattering is the most powerful technique for measuring vibrational excitation spectra. With single-crystal samples, the complete phonon dispersion relation can be measured. At small phonon wavelengths, the slope of the dispersion curve is related directly to the elastic constants of the material. Information from shorter wavelength phonons can be used to determine parameters for interatomic force models or compared to first-principles calculations. Model parameters can then be used to calculate phonon densities-of-states and consequently connect to the thermodynamics. Polycrystalline samples yield less information, but the data is directly related to the phonon density-of-states for Bravais lattices. Measurements on polycrystalline samples have the advantage of being fast and are more useful for parametric studies of the lattice dynamics (with temperature, for example). These ideas will be used to discuss several interesting systems, most notably Pu metal and Fe-Ga magnetostrictive alloys. [This work is supported by the U.S. Department of Energy.]

1:25

1pPAb2. Elastic constants, Blackman diagrams, and new lanthanide-actinide insights. Hassel Ledbetter and Albert Migliori (Los Alamos Natl. Lab., Los Alamos, NM 87545, hledbet@lanl.gov)

Using monocrystal elastic constants and Blackman diagrams, one can infer material interconnections, physical-property trends, and knowledge about interatomic bonding. After reviewing briefly the fcc-metal case, focus shifts to Ce, Pu, Th, with known monocrystal fcc elastic constants, the C_{ij}. For fcc U and a few other cases, one can deduce the C_{ij}. From these results about the actinides, there follow several conclusions: (1) elastic anisotropy ranges widely; (2) Poisson ratios are low; (3) interatomic bonding varies widely; (4) all show unusual *negative* Cauchy pressures; (5) a strong covalent-bonding component occurs, perhaps related to localized f-electrons; (6) δ-Pu behaves most oddly; (7) δ-Pu shares some bonding features with fcc alkali metals. [Work supported by DOE, NSF, State of Florida.]

1:45

1pPAb3. Elastic response of exotic ferromagnets? Veerle Keppens, Yanbing Luan, Sriparna Bhattacharya, Ivan Sergienko (The Univ. of Tennessee, Knoxville TN 37934), David Mandrus, and Brian Sales (Oak Ridge Natl. Lab., Oak Ridge, TN 37831)

Resonant ultrasound spectroscopy (RUS) measurements have been carried out as a function of temperature (2–300 K) and magnetic field (0–2 T) for a series of exotic ferromagnetic materials, including $Eu_8Ga_{16}Ge_{30}$ ($T_c \simeq 32$ K), $EuFe_4Sb_{12}$ ($T_c \simeq 82$ K), Au_4V ($T_c \simeq 46$ K), and $Yb_{14}MnSb_{11}$ ($T_c \simeq 53$ K). The different materials show remarkable differences in their elastic response near the Curie temperature, revealing a sudden softening at T_c in $EuFe_4Sb_{12}$, and a stiffening of the elastic moduli at T_c in $Yb_{14}MnSb_{11}$. Ongoing attempts to theoretically model the data using Landau theory will be discussed. [Work was supported by the National Science Foundation under Grant No. DMR0506292. Oak Ridge National Laboratory is managed by UT-Batelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.]

2:05

1pPAb4. Measuring elastic constants of novel materials and thin films using small sample resonant ultrasound spectroscopy. J. D. Maynard (The Penn State Univ., University Park, PA 16802)

One of the most important undertakings for materials is the measurement of the elastic constants. As derivatives of the free energy with respect to atomic displacements, the elastic constants are closely connected to the thermodynamic properties of the material, including the specific heat, the Debye temperature, the Gruneisen parameter, etc. The elastic constants are a sensitive probe of the lattice environment in which all solid state phenomena occur, particularly in the vicinity of a phase transition. Novel materials to which RUS might be applied are often available only in small samples, and for these cases we have developed a small sample RUS

3033   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3033**

apparatus that can measure bulk samples only a few hundred microns in size and thin films a few hundred nanometers in thickness. Thin films are particularly interesting because substrate induced strains in the lattice may produce dramatic changes in the elastic and thermodynamic properties. In this talk, small sample RUS measurements on the new superconductor magnesium diboride and colossal magnetoresistance materials, in bulk and thin film forms, will be discussed.

2:25

1pPAb5. Ultrasonic studies of hydrogen absorbing materials: Quasicrystals, nanocrystals, and intermetallic compounds. Robert G. Leisure (Dept. of Phys., Colorado State Univ., Fort Collins, CO 80523-1875)

Resonant ultrasound spectroscopy was used to study several materials—nanocrystalline palladium, quasicrystalline Ti-Zr-Ni, and a number of intermetallic compounds—that have the ability to absorb considerable amount of hydrogen, and thus are of potential use for hydrogen storage. Elastic energies are relevant to the thermodynamics of such systems, which are often described by a lattice gas model. Elastic interactions between dissolved H atoms lead to a critical temperature and precipitation of a hydride phase below the critical temperature. The energetics of hydride precipitation depend on the elastic constants. Thus, it is important that the elastic properties of such systems be measured. In addition, these materials are of interest for several other reasons. The elastic constants and internal friction for the hydrogen-free materials were measured over a temperature range of 3 to 300 K, and in a few cases to higher temperatures. These results will be discussed and interpreted. [This work was supported by the U.S. National Science Foundation under Grant DMR-0070808.]

2:45

1pPAb6. Osmium's full elastic tensor between 5 K and 300 K. Cristian Pantea (Los Alamos Natl. Lab., Los Alamos, NM 87545)

Reported values of the zero-pressure bulk modulus of Os are both above and below that of diamond, creating a controversy as to which is the least compressible material. This work is another such study, but using two measurement techniques on two types of Os specimens. The full elastic tensor for osmium was measured using resonant ultrasound spectroscopy between 5 K and 300 K on polycrystalline and monocrystalline specimens and confirmed at ambient temperature using a pulse-echo-overlap measurement. It was found that osmium is more compressible than diamond (bulk modulus: 405 GPa for Os, versus 442 GPa for diamond). The independent elastic moduli of osmium are in relatively good agreement with the theory of Fast [Phys. Rev. B **51**, 17431 (1995)]. However, the Debye temperature determined from low-temperature monocrystal elastic-moduli measurements extrapolated to 0 K disagrees with most previous reports and is in good agreement with one specific-heat measurement.

3:05−3:20   Break

Contributed Papers

3:20

1pPAb7. Excitation of elastic and capillary waves by electric field tweezers and its application to material characterization. Maiko Hosoda, Hideo Ogawa, and Hiroyasu Nomura (Tokyo Denki Univ., Ishizaka, Hatoyama, Saitama 350-0394, Japan)

We recently developed the electric tweezers system to excite the surface vibration for the measurement of local mechanical properties of soft condensed materials. A periodical electric field applied onto the material surface by a conductive thin needle induces the Maxwell stress and picks up the sample surface. We can determine the shear elasticity of the sample from the absolute displacement of the deformation, and the viscosity from its dynamic response, respectively. The numerical simulation predicts that the electric voltage of 100 V applied to the gap of 100 Em between the needle edge and surface induces the deformation of several nm, which was experimentally confirmed by the optical lever technique. The remarkable feature of the system is that the characteristic scale of the surface deformation, such as the depth and width of the medium motion, is dominated by the gap and, therefore, we can uniquely determine the complex mechanical impedance and the surface energy by observing the dynamic surface response in a wide range of the gap and oscillation frequency. The system was applied to the gelatin gel at various concentrations and we found the results obtained are well explained by the previously known complex elasticity of the material.

3:35

1pPAb8. Direct observation of phonon relaxation by optical beating light scattering spectroscopy. Keiji Sakai (Inst. of Industrial Sci., Univ. of Tokyo, 4-6-1 Komaba, Meguro-Ku, Tokyo 153-8505, Japan)

A novel dynamic light scattering system to observe elastic relaxation phenomena with hyper frequency resolution was developed. The principle of the measurement is based on the theory, which describes the dynamic structure factor of fluid under the condition of the frequency-dependent compressibility. The dynamic structure factor, which is usually composed of the Brillouin and Rayleigh triplet, is modulated and shows an additional central component that directly reflects the whole aspect of the relaxation. In the experiment, the output from a frequency-doubled cw-YAG laser was incident into the liquid sample and the power spectrum of the light scattered into the backward direction was analyzed by the optical beating spectroscopy technique. The sample is liquid acetic acid that is known to show a strong ultrasonic relaxation around 1 MHz due to the molecular association process. We could find, in the observed spectrum, the central component introduced by the phenomenon, whose relaxation frequency and strength can be obtained from the width and the intensity of the observed central peak, respectively. The results show very good agreement with those previously obtained by the conventional ultrasonic spectroscopy technique. The recent experiment of the thermal phonon resonance will also be shown.

3:50

1pPAb9. High resolution Brillouin scattering in solid and gases. Takeshi Yogi and Keiji Sakai (Dept. of Appl. Phys., Univ. of Tokyo, 4-6-1 Komaba Meguro-ku, Tokyo, 153-8505, Japan, yogitake@iis.u-tokyo.ac.jp)

The experimental system of our optical beating Brillouin spectroscopy was improved, which can observe the phonon propagation with the resolution of 1 kHz. The increased sensitivity and frequency resolution were demonstrated in the observation of phonon spectrum in solids of fused silica and also in gases, nitrogen, and freon $CHClF_2$: for their weak light-scattering ability, these substances are very tough specimens for Brillouin scattering studies, and the classical spectrometer of Fabry-Pérot etalon has been so far used for the frequency analysis. The phonon peaks were observed with the present light beating system in nitrogen and freon near 300 K over the wave-number range from $2\pi = 6.20 \times 10^3$ to 3.09×10^5 m^{-1}, the pressure range from 14.8 to 177 kPa. The observed spectra were well fitted with the Lorentzian curve representing the phonon lifetime, and the phase velocity. The phase velocity of phonon in fused silica agrees well

with that obtained by ultrasonic measurement, while the phonon dispersions obtained for gases show the relaxational behavior due to the coupling to the rotational and vibrational motion of molecules.

4:05

1pPAb10. Hypersonic phononic crystals. Taras Gorishnyy, Chaitanya K. Ullal, Martin Maldovan, Edwin L. Thomas (Dept. of Mater. Sci. and Eng., MIT, Cambridge, MA 02139), and Ji-Hyun Jang (MIT, Cambridge, MA 02139)

Phononic crystals are structures with spatially periodic variations in density and sound velocities. The most fascinating feature of phononic crystals is the existence of stop bands or band gaps in their dispersion relation, where no phonons are allowed to propagate. As a result, phononic crystals have great promise for achieving precise control over propagation of sound and other mechanical waves. In this talk our recent investigation of 2-D and 3-D hypersonic phononic crystals with band gaps in GHz frequency range will be presented and their potential applications as acoustic mirrors, lenses, waveguides, etc. will be discussed. High-quality single-crystalline structures were fabricated using interference lithography. Their phonon dispersion relation was directly measured with Brillouin light scattering. Finite-element analysis was employed to compute theoretical dispersion relation and provide interpretation for the experimentally observed propagation modes [T. Gorishnyy *et al.*, Phys. Rev. Let. **94**, 115501 (2005); T. Gorishnyy *et al.*, Physics World **18** (12), 24–29 (2005)]. [Work supported by NSF Grants DMR-0308133 and DMR-0414974.]

4:20

1pPAb11. Acoustical properties of two-dimensional sonic crystals. Francisco Cervera, Elisabeth Schwarz, Daniel Torrent, Suitberto Cabrera, and José Sánchez-Dehesa (Wave Phenomena Group, Nanophotonics Technol. Ctr., Polytechnic Univ. of Valencia, C/ Camino de vera s/n, E-46022 Valencia, Spain)

The transmittance and reflectance spectra of two-dimensional sonic crystals made of hexagonal arrays of cylindrical bars in water is reported. The balance of sound fluxes is analyzed by studying angle-resolved spectra of transmitted and reflected sound for all directions in space. A characterization of the acoustic band structure along the two high symmetry directions in the lattice (i.e., ΓX and ΓJ directions) in a large region of frequencies will be presented. It is shown that, for example, the shape of the stop band features in different types of acoustical spectra is interlinked and must be studied jointly to understand acoustical phenomena in these materials. Moreover, the dependence of acoustical properties as a function of the number of layers is also studied. The experiments are compared with theoretical simulations based on multiple scattering theory. [Work supported by MEC of Spain.]

4:35

1pPAb12. Design of resonant ultrasound systems for use with very small samples. Ben Migliori (UCSD, 9500 Gilman Dr., San Diego, CA 92122), Albert Migliori, and Jon Betts (NHMFL-LANL, Los Alamos, NM 87545)

The design of resonant ultrasound (RUS) systems for small samples, whether used in small-bore cryostats at low temperature and high field or at room temperature, presents several unique challenges. These include alignment and containment of the sample, thermal contraction of construction materials, ease of use, and robustness in the presence of large ($¿10$ T) magnetic fields. Such problems may be resolved through the use of novel (and often easily obtained) materials as well as unique designs. At the National High magnetic Field Laboratory (NHMFL) we have addressed these problems and built a sub-4 K cryostat RUS measurement system capable of reliably measuring samples '1 mm in diameter. Examples of such a cryostat systems, designed for conditions ranging from room temperature and zero field to '4 K and ¿10 T, as well as explanations of the necessary design parameters and recommended materials are presented. Examples of RUS measurements with this system are presented.

4:50

1pPAb13. Nondiffractive propagation in sonic crystals. Victor J. Sanchez-Morcillo, Isabel Perez-Arjona, Javier Redondo, Victor Espinosa (Dept. de Fisica Aplicada, Universidad Politecnica de Valencia, Ctra. Natzaret-Oliva s/n 46730 Grao de Gandia, Spain), and Kestutis Staliunas (Universitat Politecnica de Catalunya, E-08222 Terrassa, Spain)

Sonic crystals are media with a periodic modulation of the acoustic parameters, such as the density and the bulk modulus. They have recently attracted a great interest, because of their potential applications in the control of sound propagation, used as reflectors, focusers, or waveguides. All these properties are related with the dispersion introduced by the crystal anisotropy. We report on the nondiffractive propagation of sound in two-dimensional sonic crystals. It is shown that, for given frequencies and directions of incidence, a narrow sonic beam can propagate without diffractive broadening. Such nondiffractive sonic beams exist in crystals with perfect symmetry and do not require the presence of defects, differently from other waveguiding phenomena reported previously. The cancellation of diffraction has been predicted using the plane-wave expansion method to evaluate the dispersion surfaces of the crystal and the spatial dispersion (isofrequency) curves. It occurs for frequencies and wave vectors for which dispersion curves have zero curvature, denoting a transition between focusing and defocusing regimes. By means of perturbative techniques, a simple analytical expression for the nondiffractive conditions has been obtained. The phenomenon is also demonstrated by numerical integration of the acoustic equations using FDTD techniques. [Support from Spanish MEC, Project FIS2005-07931-C03-01/03, is acknowledged.]

5:05

1pPAb14. Strobo-photoelastic visualization of Lamb-type waves with negative group velocity in a solid/fluid/solid trilayer. Ken Yamamoto (Kobayasi Inst. of Physical Res., 3-20-41 Higashi-Motomachi, Kokubunji, Tokyo, 185-0022, ken-yamamoto@kobayasi-riken.or.jp)

An elastic wave with negative group velocity transports energy in the direction opposite to the phase velocity. This phenomenon exists in certain modes of Lamb waves propagating in a solid plate, but only over a very narrow range of frequency-thickness products and Poisson's ratios of a plate material. A minimum value of the negative group velocity of a mode is dominated by the Poisson's ratio of a material. A similar phenomenon also appears in Lamb-type waves that are guided in a symmetric solid/fluid/solid trilayer. Derivation of the dispersion equation of Lamb-type waves emphasizes the existence of the negative group velocity in certain modes. The existence of negative group velocity depends not only on the modes, frequency-thickness product, and Poisson's ratio, but also on the fluid layer thickness. The magnitude of negative group velocity is controlled by changing the fluid layer thickness. Direct evidence is presented of the negative group velocity that is obtained by means of stroboscopic photoelastic visualization of pulsed Lamb-type waves propagating in a glass/water/glass trilayer. A time-sequence photoelastic image shows the wave-phase behavior and the pulse envelope propagating in the opposite direction.

Session 1pSC

Speech Communication: Synthesis and Recognition of Speech (Poster Session)

Om D. Deshmukh, Cochair

Univ. of Maryland, Electrical and Computer Engineering Dept., A.V. Williams Bldg., College Park, MD 20742

Koichi Shinoda, Cochair

Tokyo Inst. of Technology, Graduate School of Information Science and Engineering, Dept. of Computer Science, 2-12-1 W8-81, Ookayama, Meguro-ku, Tokyo 152-8552, Japan

Contributed Papers

All posters will be on display from 2:00 p.m. to 6:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 2:00 p.m. to 4:00 p.m. and contributors of even-numbered papers will be at their posters from 4:00 p.m. to 6:00 p.m.

1pSC1. Improving quality of small body transmitted ordinary speech with statistical voice conversion. Hidehiko Sekimoto, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Inst. of Sci. and Technol., 8916-5, Takayama-cho, Ikoma-shi, Nara, 630-0192 Japan)

The explosive spread of cellular phones enables us to communicate with each other at any time or place. Although cellular phones are convenient, there are still some problems. For example, it is difficult to send intelligible speech under noisy conditions, which is a fatal problem especially when talking privately using small speech in crowds. To improve the quality of small speech under such situations, a new speech communication style is proposed using a nonaudible murmur (NAM) microphone [Nakajima *et al.*, Eurospeech (2003), pp. IV-2601–2604]. The NAM microphone is robust to eternal noise, although body transmission causes quality degradation. In this paper, to improve the sound quality of small body transmitted ordinary speech (SBTOS), which is small speech recorded with a NAM microphone, two conversion methods that reflect a statistical voice conversion method based on Gaussian mixture models (GMMs) [Toda *et al.* Interspeech (2005), pp. 1957–1960] are conducted. One conversion method is from SBTOS to ordinary speech (SBTPS-to-SP), and the other is from SBTOS to small speech (SBTOS-to-SSP). SBTOS-to-SSP has more consistent correspondence of voiced/unvoiced segments between input and output speech than SBTOS-to-SP. The results of objective and subjective evaluations show that SBTOS-to-SSP outperforms SBTOS-to-SP.

1pSC2. Evaluation of eigenvoice conversion based on Gaussian mixture model. Yamato Ohtani, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Grad. School of Information Sci., Nara Inst. of Sci. 8916-5, Takayama-cho, Ikoma-city, Nara, Japan 630-0192)

Eigenvoice conversion (EVC) has been proposed as a new framework of voice conversion (VC) based on the Gaussian mixture model (GMM) [Toda *et al.*, "Eigenvoice Conversion Based on Gaussian Mixture Model," ICSLP, Pittsburgh, Sept. 2006]. This paper evaluates the performance of EVC in conversion from one source speaker's voice to an arbitrary target speakers' voices. This framework trains canonical GMM (EV-GMM) in advance using multiple parallel data sets consisting of utterance pairs of the source and many prestored target speakers. This model is adapted to a specific target speaker by estimating a small number of free parameters using a few utterances of the target speaker. This paper compares spectral distortion between converted and target voices in EVC with conventional VC based on GMM when varying the amount of training data and the number of mixtures. Results show EVC outperforms conventional VC when using small amounts of training data. EVC can effectively train a

complex conversion model using the information of many prestored speakers. By contrast, conventional VC needs a large-sized parallel data set for training. It also shows the results of subjective evaluations of speech quality and conversion accuracy for speaker individuality.

1pSC3. Evaluating naturalness of speeches morphed by independently using the interpolation ratios of the time-frequency axes and amplitude. Toru Takahashi, Masanori Morise, and Toshio Irino (Faculty of Systems Eng., Wakayama Univ., 930, Sakaedani, Wakayama, Japan, 640-8510, tall@sys.wakayama-u.ac.jp)

In speech morphing procedure [Kawahara *et al.*, Proc. IEEE-ICASSP 2003, Vol. I, pp. 256–259], two parameters exist for controlling the degree of morphing between the source and target: the interpolation ratios of the time-frequency axes and amplitude. Conventional morphing methods use only a single path in which both ratios are identical. It has, however, been reported that perception of naturalness degrades around a morphing ratio of 0.5. It was assumed that there would be better combination of the ratios with maintaining the quality. It was surveyed that the naturalness contour is within a two-dimensional morphing space. The morphed speech sounds are synthesized in combinations of 11 axis ratios (between 0 and 1 by 0.1 step) and 11 amplitude ratios. Synthetic sounds (121 in all) were presented to subjects to judge the naturalness rate. The optimum path will be described at a presentation to be given at this conference. [This research was partly supported by the "e-Society Leading Project" of the Ministry Education, Culture, Sports, Science, and Technology.]

1pSC4. Building an English speech synthetic voice using a voice transformation model from a Japanese male voice. Akemi Iida (School of Media Sci., Tokyo Univ. of Technol., 1404-1, Katakura-cho, Hachiouji, Tokyo, 192-0982, Japan, ake@media.teu.ac.jp), Shimpei Kajima, Kiichi Yasu, Takayuki Arai (Sophia Univ., Tokyo, 102-8554, Japan), and Tsutomu Sugawara (Sophia Univ., Tokyo, 102-8554, Japan)

This work reports development of an English speech synthetic voice using a voice transformation model for a Japanese amyotrophic lateral sclerosis patient as part of a project of developing a bilingual communication aid for this patient. The patient, who had a tracheotomy 3 years ago and had difficulty in speaking, wishes to speak in his own voice in his native language and in English. A Japanese speech synthesis system was developed using ATR CHATR 6 years ago and the authors have worked in developing a diphone-based synthesis using FESTIVAL speech synthesis system and FESTVOX by having the patient read the diphone list. However,

3036   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3036**

it was not an easy task for the patient to phonate and, moreover, to pronounce words in a foreign language. We therefore used a voice transformation model in FESTIVAL to develop the patient's English speech synthetic voice which enables text-to-speech synthesis. We trained using 30 sentences read by the patient and those synthesized with an existing FESTIVAL diphone voice created from a recording of a native English speaker. An evaluation including a listening experiment was conducted and the result of this voice conversion showed that the synthesized voice successfully reflected the voice quality of the patient.

1pSC5. An MRI-based time-domain speech synthesis system. Tatsuya Kitamura, Hironori Takemoto, Parham Mokhtari (NICT/ATR Cognit. Information Sci. Labs., 2-2-2 Hikaridai, Seikacho, Sorakugun, Kyoto, 619-0288, Japan), and Toshio Hirai (Arcadia, Inc., Minoushi, Osaka, 562-0003, Japan)

A speech synthesis system was developed based on Maedas method [S. Maeda, Speech Commun. **1**, 199–229 (1982)], which simulates acoustic wave propagation in the vocal tract in the time domain. This system has a GUI interface that allows fine control of synthesis parameters and timing. In addition, the piriform fossae were included to the vocal tract model, resulting in antiresonances in speech spectra at the frequency region from 4 to 5 kHz. The system can produce all the Japanese phonemes using vocal tract area-functions (VTAFs) extracted from 3-D cine-MRI obtained during production of VCV or CVCV sequences for a male speaker. The system can be used to synthesize Japanese sentences with high naturalness and intelligibility by concatenating segmental units and controlling the glottal source using the GUI interface. Since a time-varying VTAF is obtained by interpolating between VTAFs, the dataset size of the system is significantly smaller than that of corpus-based speech synthesizers. The speaker-specific VTAFs and inclusion of the piriform fossae permit us to reproduce speaker-specific spectral shapes, not only the lower formants but also higher frequency regions that contribute to the perception of speaker individualities. [Work supported by NICT, SCOPE-R, and Grant-in-And for Scientific Research of Japan.]

1pSC6. Database size and naturalness in concatenative speech synthesis. H. Timothy Bunnell, James T. Mantell, and James B. Polikoff (Speech Res. Lab., A. I. duPont Hospital for Children, 1600 Rockland Rd., Wilmington, DE 19803)

Unit concatenation TTS systems seek to maximize perceived naturalness by minimizing the amount of signal processing applied to the recorded speech on which they are based. To generate distinct suprasegmentals for a given segmental sequence (e.g., to convey variation in focus or emotion), it is necessary to record and store multiple instances of the same segments that vary in fundamental frequency and voice quality. At the expense of naturalness, concatenative systems can store a minimal segmental inventory and synthesize suprasegmental factors by manipulating $f0$ and voice quality via signal processing. Classic diphone synthesis (where only a single instance of each diphone sequence is stored) represents the limiting case of this strategy. The present study explores aspects of the trade-off between perceived naturalness and segmental inventory size using the ModelTalker TTS system. Twenty-five speakers each recorded about 1650 utterances. From these, databases were constructed that limited the maximum number of alternate diphones usable for synthesis in five conditions to 1, 5, 10, 20, and 40. Sentences were synthesized from these databases using either full or no $f0$ control. Results of listening tests wherein subjects rate the naturalness of each sentence will be presented. [Work supported by NIDCD.]

1pSC7. Optimization of target cost weights in concatenative speech synthesis with very short segments of 5-ms duration. Toshio Hirai (Arcadia, Inc., 3-1-15 Nishishoji, Mino, Osaka 5620003, Japan, thirai@arcadia.co.jp)

If a concatenative speech synthesis system uses more short speech segments, it increases the potential to generate natural speech because the concatenation variation becomes greater. Recently, a synthesis approach was proposed in which very short (5 ms) segments are used [T. Hirai and S. Tenpaku, 5th ISCA Speech Synthesis Workshop, pp. 37–42 (2004), http://www.ssw5.org/]. In that approach, the target cost (how close a database segment is to desired segment) was defined as the simple average of the root mean squares (rms's) of the difference between the features of the database segment and the desired segment for simplicity. Therefore, it has been expected to optimize the weight of each rms. A Japanese speech database was used to optimize the weights and to evaluate its effects. For corpus construction, 150 utterances were selected from the database. Ten other utterances were selected randomly for generation of feature time series as natural targets for synthesis. Half of them were used for optimization using many weight combinations in synthesis to determine the optimal weight set that shows the minimum concatenation distortion. Distortion for the other half of the utterance synthesis with optimized weights was reduced 34.4% compared to the former approach.

1pSC8. Improving Japanese syllable unit selection using search space with small acoustical features variance. Takaaki Moriyama and Seiichi Tenpaku (Arcadia, Inc., 3-1-15, Nishishoji, Mino, Osaka, 5620003 Japan, taka@arcadia.co.jp)

Usually in concatenative text-to-speech systems, suitable units are selected according to acoustical features. However, a method to reduce calculation costs has been proposed, which is using syllable notations and positions along with the mora length [Murakami *et al.*, Trans. IEICE **J85-D-II**(7), 1157–1165 (2002)]. An enlargement of the corpus to synthesize longer phrases lowers the quality of synthesized speech because of increased feature variance within syllable units with identical notation. Methods to produce detailed groups in which the feature's variance is smaller than syllable units, and to synthesize speech using these groups, are proposed. To produce the groups, a mel-cepstrum distance is used as a measure for evaluation. In this method, the quality of synthesized speech that is given an ideal group sequence might improve, as the distance between units is close within the same group. As a first step to evaluate the proposed method, five groups were generated from a Japanese corpus; then, speech that was synthesized to replace one unit in natural speech to others in the same groups was examined. Results of this listening experiment indicated that an average of the naturalness for these speech types was greater than 80%.

1pSC9. Toward hidden Markov model-based spontaneous speech synthesis. Tatsuya Akagawa, Koji Iwano, and Sadaoki Furui (Dept. of Comput. Sci., Tokyo Inst. of Technolology, 2-12-1-W8-77 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan, akagawa@furui.cs.titech.ac.jp)

This paper investigates several issues of spontaneous speech synthesis. Although state-of-the-art synthesis systems can achieve highly intelligible speech, their naturalness is still low. Therefore, much work must still be done to achieve the goal of synthesizing natural, spontaneous speech. To model spontaneous speech using a limited amount of data, we used an HMM-based speech synthesizer based on three features: cepstral features modeled by HMMs, and duration and fundamental frequency features modeled using Quantification Theory Type I. The models were trained with approximately 17 min of spontaneous lecture speech, from a single speaker, which was extracted from the Corpus of Spontaneous Japanese (CSJ). For comparison, utterances by the same speaker, reading a transcription of the same lecture, were used to train analogous models for read speech. Spontaneity of the synthesized speech was evaluated by subjective pair comparison tests. Results obtained from 18 subjects showed that the preference score for the synthesized spontaneous speech was significantly

1p TUE. PM

higher than that for the synthesized read speech. This implies that HMM-based speech synthesis using actual spontaneous utterances for model training is effective at producing natural speech. Additional subjective evaluation tests were also conducted to analyze the effects of individual features on the impression of spontaneity.

1pSC10. Clustering strategies for hidden Markov model-based polyglot synthesis. Javier Latorre, Koji Iwano, and Sadaoki Furui (Tokyo Inst. of Technol., Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552, Japan, latorre@furui.cs.titech.ac.jp)

A speaker-adaptable polyglot synthesizer is a system that can synthesize multiple languages with different voices. In our approach, we mix data from multiple speakers in different languages to create an HMM-based multilingual average voice. This average voice is then adapted to a target speaker with a small amount of speech data from that speaker. Using this adapted voice our system is able to synthesize any of the languages for which it was trained, with a voice that mimics that of the target speaker. The main technical difficulty of this approach is how to cluster the acoustic models to create a voice that sounds equal for all the languages. In this paper we analyze different possible strategies for clustering the HMMs and compared them on a perceptual test. The results showed that questions that refer only to the phonetic features and ignore the original language of the sounds produced voices with more foreign accent but that sound clearly more consistent across languages. Surprisingly, this foreign accent affected neither the intelligibility nor the naturalness of the synthetic speech. Therefore for polyglot speech synthesis the overall better performance was obtained by clustering strategies based only on the phonetic features of the sounds.

1pSC11. Concept-to-speech conversion for reply speech generation in a spoken dialogue system for road guidance and its prosodic control. Yuji Yagi, Seiya Takada, Keikichi Hirose, and Nobuaki Minematsu (School of Eng., Univ. of Tokyo, Bldg. No. 2, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan, yagi@gavo.t.u-tokyo.ac.jp)

Because of advancements of speech and language processing, a number of spoken dialogue systems have been constructed. However, because most of them adopt existing text-to-speech synthesizers to generate output speech, it is rather difficult to reflect all the linguistic information that is obtained during the reply sentence generation. To resolve this situation, a framework must correctly reflect higher-level linguistic information, such as syntactic structure and discourse information, on the prosody of output speech: concept-to-speech conversion, where reply sentences are generated from information (to be transmitted) and converted into speech in a unified process. A spoken dialogue system for road guidance was constructed, and concept-to-speech conversion was realized in the system. The linguistic information of the generated sentence is handled in tag LISP form to retain the syntactic structures throughout the process. Moreover, a new method of sentence generation from concept was realized with this system: it handles a concept in phrase units and aggregates them to form a sentence. It is tested whether the linguistic information could be reflected properly on the prosody of output speech. Results of listening experiments verified the effectiveness of our proposed method.

1pSC12. Vector-based language generation for associatively evoking environmental images. Kiyoaki Aikawa and Akemi Iida (School of Media Sci., Tokyo Univ. of Technol., 1404-1 Katakuracho, Hachioji, Tokyo 192-0982, Japan)

A novel language generation method was proposed for spoken dialogue systems. The method selected verbal expressions for sensory inputs using a vector-based method. A weather-reporter system was implemented with seven input categories of weather parameters including temperature, humidity, wind velocity, and cloud cover. The system selected the most appropriate short phrase for representing the measured weather parameters

to let the listener associatively imagine the reporter's weather condition descriptions. The system included 149 short phrases for weather expressions, each represented by a 62-element vector called the template vector. Given the 62-element input vector for a set of weather parameters, one or two phrases were selected based on the similarity measured between the input and template vectors. The system was able to produce an expression by combining a couple of phrases. A dissimilarity measuring algorithm was proposed to assign the conjunction "and" or adversative "but" automatically to conjoin two phrases. A bilingual capability was also added for enabling the system to output English and Japanese phrases. For English outputs, sentence-initial phrases were added to the selected phrases depending on the similarity value, while sentence-finals were added for Japanese outputs. Evaluation experiments showed that automatic conjunction assignment can function with high reliability.

1pSC13. Development project for screen reader interface dynamically hastens speech while giving emphasized information to the tactile sense. Tohru Ifukube (Res. Ctr. for Adv. Sci. and Technol., The Univ. of Tokyo, Tokyo, 153-8904, Japan) and Shinta Kimura (Animo Ltd., Naka-ku, Yokohama-shi, 231-0015, Japan)

In present screen readers for blind people, users are still required to continue listening to synthesized voices and also they might easily miss important points of a document. A development project of a new screen reader interface has been promoted in cooperation with Japan's Ministry (METI). The interface named TAJODA can dynamically control the speech rate while giving rich texts to the tactile sense. The speech rate can be changed according to word units by clicking a button of the interface using a thumb. From evaluation tests using sightless people, the maximum speech rate was determined at around 2000 morae/minute. Seven vibro-tactile patterns were selected for presenting rich texts onto an index finger using a tactile matrix display (2×8). Both speech and tactile information are synchronized automatically through a USB interface. It was also ascertained that users can read some documents two to three times faster using the TAJODA interface [Asakwa *et al.*, IEICE Trans. **E87**-D(6) (2004)].

1pSC14. Decision-tree-based $F0$ quantization for hidden Markov model-based speech coding at 100 bit/s. Yoshihiro Itogawa, Heiga Zen, Yoshihiko Nankaku, Akinobu Li, and Keiichi Tokuda (Dept. of Comput. Sci., Nagoya Inst. of Technol., Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan)

A decision-tree-based quantization scheme for a very low bit rate speech coder based on HMMs is described. The encoder carries out HMM-based phoneme recognition and then recognized phonemes, state durations, and $F0$ sequence are quantized, Huffman coded, and transmitted. In the decoder, sequences of mel-cepstral coefficient vectors and $F0$'s are generated from the concatenated HMM-using the HMM-based speech synthesis technique. Finally, a speech waveform is synthesized by the MLSA filter using the generated mel-cepstral coefficient and $F0$ sequences. In the previous system, we train an MSD-VQ codebook for each phoneme for $F0$ quantization. Although this scheme can quantize $F0$ sequences efficiently, to achieve a better speech quality, larger codebook sizes are required. It leads to an increase in the bit rate of the system. To avoid this problem, we cluster $F0$ sequences using phonetic decision trees and then train a codebook for each leaf node. In the encoding and decoding, codebooks to be used can be determined by tracing the decision tree. It allows us to use smaller codebook sizes since the number of codebooks can be augmented without increase in bit rate. A subjective listening test result shows that the proposed scheme improves the quality of coded speech.

1pSC15. Speech enhancement during analysis-synthesis by harmonic plus noise model. Parveen K. Lehana and Prem C. Pandey (EE Dept., Indian Inst. of Technol. Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in)

Speech analysis-synthesis in codecs and voice transformation often introduces distortion in the resynthesized output, particularly if the input is corrupted by noise. The harmonic plus noise model (HNM), particularly suited for voice transformation, divides speech signal into harmonic and noise parts. The harmonic part is synthesized by taking the summation of pitch harmonics up to a frequency called maximum voiced frequency, and the spectrum above this frequency is attributed to the noise part. Removal of nonharmonic components in harmonic part may result in noise rejection, while the errors in the estimation of harmonics may lead to distortion. Investigations were conducted to assess the speech enhancement obtainable through HNM on broadband noise corrupted speech signal. Listening tests showed that both the input and the resynthesized signals were perceived as distortion-free for SNR greater than 10 dB. In the 0–10-dB range, both the signals had no loss of intelligibility, but listeners rated the resynthesized signal as having higher quality. No enhancement was obtained for SNR values lower than 0 dB. Hence, for higher noise levels, spectral subtraction or other enhancement techniques should be employed along with analysis-synthesis.

1pSC16. Enhancement of electrolaryngeal speech by spectral subtraction with minimum statistics-based noise estimation. Priyanko Mitra and Prem C. Pandey (EE Dept., Indian Inst. of Technol. Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in)

Transcervical electrolarynx is a vibrator held against the neck tissue by a laryngectomy patient to provide excitation to the vocal tract, as a substitute to that provided by glottal vibrations. The electrolaryngeal speech suffers from lack of voicing and pitch control, low-frequency deficiency, and presence of background noise, caused by leakage of acoustic energy from the vibrator and vibrator-tissue interface. Pitch-synchronous application of spectral subtraction has been used for reducing the self-leakage noise, with averaging-based noise estimation on an initial segment with closed lips. But, the leakage noise spectrum varies with speech production and vibrator orientation and pressure. A dynamic estimation of noise is carried out by using minimum statistics, with two-pitch period analysis frames and one period overlap. Minimum value of each spectral sample in a set of past frames is used for dynamically estimating the magnitude spectrum of noise. This is subtracted from the magnitude spectrum of noisy speech and the signal is reconstructed using the original phase spectrum. Smoothing of the estimated noise spectrum resulted in better noise cancellation. Quality of electrolaryngeal speech is further enhanced by spectral deficit compensation of the magnitude spectrum.

1pSC17. Evaluating the perceptual quality of speech signals enhanced using the Modified Phase Opponency model. Om D. Deshmukh and Carol Y. Espy-Wilson (Dept. of Elec. & Comput. Eng., A. V. Williams Bldg., Univ. of Maryland, College Park, MD 20742)

Previously, we presented a speech enhancement algorithm based on the auditorily motivated MPO model. It was shown that the MPO-based speech enhancement scheme is robust to fluctuating noise. Compared to the output of several other speech enhancement techniques, the MPO-enhanced speech signals exhibit the lowest increase in Itakura-Saito, log-area-ratio, and log-likelihood-ratio-based distortion measures as the SNR reduces. In the present work, the performance of the MPO speech enhancement is compared with that of the other techniques using subjective evaluations. Three normal-hearing subjects were recruited to perform two-alternative, forced-choice preference tasks. Each sentence was presented twice: one of them was enhanced using the MPO scheme and the other was either left unprocessed or enhanced using one of three other speech enhancement techniques. Sentences are 5-digits long and corrupted by additive subway noise at −5, 0, 5, or 15-dB SNR. Most of the time, the MPO-enhanced speech signal was preferred weakly. Moreover, the MPO-

processed clean speech signal was preferred over clean unprocessed speech signal just as many times as the clean unprocessed speech signal was preferred, indicating that the MPO processing does not introduce any perceptual artifacts in clean speech. [Work supported by NSF BCS0236707 DOD H9823005C0425.]

1pSC18. Noise reduction of speech using multiband noise subtraction and adaptive filtering methods in two stages. Md. Ekramul Hamid, Keita Ogawa, and Takeshi Fukabayashi (Grad. School of Elec. Sci. and Tech., Shizuoka Univ., 3-5-1 Jouhoku, Hamamatsu 432-856, Japan)

Real-world noise is mostly colored having nonuniform power distribution over the whole spectrum. A single-channel two-stage method with multiband processing technique is presented in this paper. Initially a noise dominant signal is estimated from the observed speech and obtained the first noise-reduced speech by subtracting the signal. The noise-dominant signal is estimated from the valleys of the spectrum whose amplitudes are tuned by the estimated noise degree. The bandwise estimated noise degree is obtained from *a priori* function as a function of a parameter obtained from the autocorrelation of the noisy speech. The outcome of the first stage is allowed to process by the second stage in bandwise to obtain more enhanced speech. Blind source separation technique is employed that consists of two adaptive filters. The least-mean-square algorithm based on steepest-descent method is implemented to update the filter coefficients in an iterative manner. The method addresses the situations in which the input signal-to-noise ratio varies substantially and a fixed number of iteration of the least-mean-square algorithm for each signal-to-noise ratio is time consuming. Therefore, a function has been proposed that estimates the minimum iteration number required for a given value of the noise degree of noisy speech.

1pSC19. Speech interface for operating information technological home appliances. Masanori Enoki, Tadashige Noguchi, Tsubasa Arai, Ayako Miyake (Doshisha Univ., 1-3, Tatara-Miyakodani, Kyotanabe, Kyoto, 610-0321, Japan), Kohei Yasui, Kenko Ota, Masuzo Yanagida (Doshisha Univ., Kyoto, 610-0321, Japan), and Masaki Ida (OMRON Corp., Kyoto, 619-0283, Japan)

Information technological (IT) home appliances are becoming highly advanced by being connected to the Internet. Conventional push-button systems on remote controllers cannot suffice for IT home appliances because of their complicated requirements for versatile and complex functions. Speech input is expected to be a good substitute or supplement for push-button systems. Effects of introducing speech input to a TV system are evaluated, implementing facilities of on-line program retrieval besides ordinary functions of controlling the TV sets. One of the problems with introducing speech input is degradation of speech recognition rate due to being affected by surrounding noises and waves reflected on the surface of walls, ceiling, and so on. So, the test system introduced reduction of known noise, dereverberation scheme, delay-and-sum beamformer and blind source separation by independent component analysis as preprocessing for speech recognition. Effects of introducing these types of preprocessing are investigated comparing speech recognition rates in case of employing these types of preprocessing with those in the case of not employing these types of preprocessing. [Work supported by Knowledge Cluster Project, MEXT, and by Academic Frontier Project, Doshisha Univ.]

1pSC20. Speech detection using Real Adaboost in car environments. Tetsuya Takiguchi, Hiroyoshi Matsuda, and Yasuo Ariki (Dept. of Comput. and System Eng., Kobe Univ., 1-1 Rokkodai, Nada, Kobe, 657-8501, Japan, takigu@kobe-u.ac.jp)

In real noisy environments, a speech detection algorithm plays an especially important role for noise reduction, speech recognition, and so on. In this paper, a speech/nonspeech detection algorithm using Real Adaboost

is described, which can achieve extremely high detection rates. Boosting is a technique of combining a set weak classifiers to form one high-performance prediction rule, and Real Adaboost [R. E. Schapire and Y. Singer, Mach Learn. **37**, 3, 297–336, (1999)] is an adaptive boosting algorithm in which the rule for combining the weak classifiers adapts to the problem and is able to yield extremely efficient classifiers. The Real Adaboost algorithm is investigated for speech/nonspeech detection problem. The proposed method shows an increasing speech detection rate in car environments, comparable with that of the detector based on GMM (Gaussian mixture model), where the detection accuracy rate was 98% for the proposed method and 92% for GMM in a car at highway speed. The results of the experiments clarified the effectiveness of the proposed method.

1pSC21. Nonrestrictive and long-term monitoring system of utterance and a laughing voice: Laughometer. Masafumi Matsumura, Yutaka Kawabata, Ryoji Suzuki, and Masaki Yoshida (Faculty of Biomed. Eng., Osaka Electro-Commun. Univ., 1130-70 Kiyotaki, Shijonawate, Osaka 575-0063, Japan)

It is said that laughter is the best medicine. Laughter is known to be effective for relieving stress and promoting health. Therefore, this study is intended to clarify the relationship between laughter and physiological factors related to health promotion. We developed a hands-free wireless system (Laughometer) that monitors laughter (a laughing voice) in every-day conversations. This Laughometer detects vocal cord vibrations using a bone-conductive microphone attached to the skin on the neck. Based on data from repeated vocalizations of laughter, this apparatus can distinguish features of laughter in everyday conversations. Based on the mean fluctuation (standard deviation) in syllable intervals, 90% of laughs are differentiable from normal conversation. While attending a live rakugo comedy show, the number and duration of laughs were monitored by analyzing laryngeal voice (approximately 1-h duration). The mean differentiation rate for laughing voice was 94%, thereby confirming the efficacy of laughter recognition based on fluctuation in syllable intervals.

1pSC22. Development of nonvoice dialogue interface for robot systems. Ryuichi Nisimura and Aki Hashizume (Faculty of Systems Eng., Wakayama Univ., 930 Sakaedani, Wakayama 640-8510, Japan, nisimura@sys.wakayama-u.ac.jp)

In this presentation, a nonvoice dialogue interface scheme for robot systems to improve the flexibility of human-robot interaction is proposed. A prototype system with a microphone in a plastic doll has been developed to examine its potential for distinguishing interaction noises and human utterances. It realizes a simple, one-input one-response interaction with perception of a user's action. For example, the system can reply "Ouch!" when its body or head is beaten by a user. It is also able to make a response to a calling by the voice. This scheme would provide a new interaction style, unlike conventional spoken dialogue systems in which a user's rough actions were treated as meaningless obstacles. It identifies the source of noises based on GMM (Gaussian mixture model) noise recognition [A. Lee *et al.*, Proc. INTERSPEECH, 1, 173–176 (2004)]. Thirteen-class GMMs were constructed from voices and noises like slapping or stroking the head, and knocking the body, which were recorded via trial testing the prototype system. It is possible to discriminate noises and utterances by comparing acoustic likelihoods from GMMs. The experiment investigates performances of acoustic feature vectors consisting of MFCC (mel frequency cepstral coefficients) to identify interaction noises.

1pSC23. An evaluation of hands-free voice activity detection algorithm based on weighted cross-power spectrum phase analysis and zero crossing detector. Takamasa Tanaka, Yuki Denda, Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, rs012019@se.ritsumei.ac.jp), and Takanobu Nishiura (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

VAD (voice activity detection) is necessary for recognizing hands-free speech. However, conventional VADs could not achieve enough detection performance in real noisy environments because they utilize only time-sequence information or only spatial-sequence information of captured signals. To cope with this problem, we have already proposed noise-robust hands-free VAD by composing the time-sequence and spatial-sequence information. The conventional ZCD (zero crossing detector) was employed to acquire the time-sequence information, and weighted CSP (cross-power spectrum phase) analysis that was a specialized DOA (direction of arrival) estimation algorithm for speech signal was also employed to acquire the spatial-sequence information. However, this algorithm can insufficiently improve the VAD performance in higher noisy environments. To overcome this problem, in this paper, we try to utilize talker direction information and attempt the environmental adaptation of our former algorithm. As a result of evaluation experiments in real acoustic environment, the proposed hands-free VAD could achieve a more effective performance than the conventional VAD and our former VAD. [Work supported by MEXT of Japan.]

1pSC24. A study of robust omnidirectional audio-visual talker localization algorithm with microphone array and omnidirectional image. Yuki Denda (Grad. School of Sci. and Eng., 1-1-1 Nojihigashi, Kusatsu, Japan, gr021052@se.ritsumei.ac.jp), Tanaknobu Nishiura, and Yoichi Yamashita (College of Information Sci. and Eng., 1-1-1 Nojihigashi, Kusatsu, Japan)

In video conferencing environments, it is very important to localize the talker. However, conventional audio signal-based algorithms often suffer from audio interference, and conventional visual signal-based algorithms fail in the presence of visual interference. To deal with these problems, this paper proposes a robust omnidirectional audio-visual talker localization algorithm that not only exploits audio feature parameters, but also subordinately uses visual feature parameters. To achieve omnidirectional audio-visual talker localization, paired-omnidirectional microphones are employed as an audio sensor, and an omnidirectional camera is employed as a visual sensor. For robust talker localization, audio feature parameters are extracted using weighted cross-power spectrum phase (CSP) analysis and CSP coefficient subtraction, and visual feature parameters are extracted using background subtraction and skin-color detection. The talker is finally located by the fusing of weighted audio/visual feature parameters, and the weight of this feature parameter fusion is automatically controlled based on the reliable criterion of audio feature parameters. The results of localization experiments in an actual room revealed that the proposed audio-visual talker localization algorithm is superior to that of conventional localizers using only audio or visual feature parameters, but not both. [Work supported by MEXT of Japan.]

1pSC25. Application of harmonic plus noise model for enhancing speaker recognition. Gidda Reddy Gangula, Prem C. Pandey, and Parveen K. Lehana (EE Dept., Indian Inst. of Technol. Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in)

Speaker recognition systems mostly employ mel frequency cepstral coefficients (MFCC). Performance of these systems is generally affected by background noise, transmission medium, etc. Further, they do not perform well in text-independent environment with limited training data. For enhanced performance, the set of parameters used should separate the speaker-dependent information from the linguistic information. Towards this end, application of parameters of the harmonic plus noise model (HNM)-based analysis is investigated. HNM divides the speech spectrum

into harmonic and noise bands separated by a dynamically varying maximum voiced frequency. This frequency is a speaker-dependent parameter and its estimation is not affected by moderate SNR degradation. A speaker recognition system was devised using three HNM parameters, namely, maximum voiced frequency, relative noise band energy, and pitch. It gave-performed comparably to that of MFCC-based recognition, for a group of ten speakers. An enhanced performance was observed by using MFCC and the three HNM parameters together, indicating the suitability of HNM for speaker recognition.

1pSC26. Noisy speech recognition using modification of spectral envelopes and smoothing and emphasis method by image filters. Yoichi Midorikawa and Masanori Akita (Faculty of Eng., Oita Univ., 700 Dannoharu, Oita, 870-1192, Japan)

The purpose of this study is to propose an improvement of the speech recognition under the noisy environment. Automatic speech recognition (ASR) systems are most effective in noiseless environments. We propose a method for ASR with signals contaminated with colored noise using modification of the spectral envelope shape. Our previous study has proposed a method of speech recognition for the signals polluted with color noise by modifying spectral envelope shape by rules. The proposed method is based on cepstral analysis. We have proposed the modified rules for adding valleys are defined. These methods are applied to the enhancement system of noisy signals and the performance of the system is improved. However, these rules only make use of frequency domain. Therefore, the cepstral coefficients of neighboring frames become independent of time domain. In this study, we apply the smoothing and emphasis method by image filters to the modification of spectral envelope shape. As a result, performance of speech recognition is improved.

1pSC27. Noisy speech recognition based on codebook normalization of discrete-mixture hidden Markov models. Tetsuo Kosaka, hidden Markov Models, Masaharu Katoh, and Masaki Kohda (Faculty of Eng., Yamagata Univ., Yonezawa-city, Yamagata, 992-8510 Japan, tkosaka@yz.yamagata-u.ac.jp)

This paper presents a new method of robust speech recognition under noisy conditions based on discrete-mixture HMMs (DMHMMs). The DM-HMMs were proposed originally to reduce calculation costs. Recently, we applied DMHMMs to noisy speech recognition and found that they were effective for modeling noisy speech [Kosaka *et al.*, Proc. of ICA04 (2004), Vol. II, pp. 1691–1694]. For further improvement of noisy speech recognition, we propose a novel normalization method for DMHMM codebooks. The codebook normalization method is based on histogram equalization (HEQ). The HEQ is commonly applied for feature space normalization. The DMHMM codebooks were normalized in this study. Therefore, this method is considered as normalization not in feature space, but in model space. Some merits are inherent in choosing model space normalization. In model space normalization, a transformation function can be prepared for each acoustic model. In addition, it is not necessary to normalize input parameters frame-by-frame. The proposed method was compared with both conventional continuous-mixture HMMs (CHMMs) and DMH-MMs. That comparison showed that the proposed method obtained the best performance: an average relative improvement of 29.2% over the CHMM baseline.

1pSC28. Robust speech recognition for home environments using factorial hidden Markov models. Agnieszka Betkowska, Koichi Shinoda, and Sadaoki Furui (Dept. of Comput. Sci., Tokyo Inst. of Technol., 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552 Japan, agabet@furui.cs.titech.ac.jp)

In this paper we present a speech recognition system for the home environment that is robust to nonstationary sudden noise such as doors slamming, knocking, and falling objects. This noise appears rapidly and

lasts for a short period of time. Additionally the SNR is difficult to estimate. As a compensation model for this task, we investigated the use of the factorial hidden Markov model (FHMM) architecture proposed by Deoras (2004). In our method, a hidden Markov Model (HMM) for each word in a dictionary and an HMM for sudden noise are employed to build a factorial HMM to recognize noisy speech. Since the factorial HMM defined by Deoras applies only to static features of the observation vector during the recognition process, we extended it to work with dynamic features. The proposed method was evaluated with a database recorded by a personal robot, PaPeRo, in the houses of 12 Japanese families. The experiments were conducted for both synthesized speech (at different SNRs) and real noisy speech, which was recorded by the robot. In both cases, our method showed significant improvement over HMMs trained with clean speech data.

1pSC29. A phoneme duration model considering speaking-rate and linguistic features for speech recognition. Yuichi Ohkawa (Grad. School of Educational Informatics, Tohoku Univ., 27-1, Kawauchi, Aoba-ku, Sendai 980-8576, Japan), Akinori Ito, Motoyuki Suzuki, and Shozo Makino (Tohoku Univ., Aramaki, Aoba-ku, Sendai 980-8579, Japan)

In this paper, we proposed a method of phoneme duration modeling for speech recognition. A phoneme with extremely short or long duration often causes a decline of performance of speech recognition. In order to improve performance of recognition, an estimation of phoneme duration determined by various parameters is required. However, there was no usual method of duration modeling for speech recognition considering the influence of both speaking-rate and linguistic feature (phoneme location in sentence, part-of-speech, *et al.*), which influence phoneme duration strongly. Therefore, we modeled influence of speaking-rate by two-dimensional normal distribution of phoneme duration and local average of vowel duration. Each normal distribution is determined by tree-based clustering with various questions, which include linguistic feature. With an experiment of estimation of phoneme duration by this model, we acquired 20.8% reduction of standard deviation of estimation error. We also used the proposed duration model for rescoring of *N*-best hypothesis of speech recognition. With an experiment of rescoring of recognition results for spontaneous speech, we acquired significant reduction of 4.7% in phoneme error rate.

1pSC30. Investigation of model-based error rate estimation techniques for speaker-independent large-vocabulary recognition tasks. Atsuhiko Kai and Shoji Nakamori (Faculty of Eng., Shizuoka Univ., 3-5-1, Johoku, Hamamatsu, Shizuoka, 432-8561, Japan)

This paper proposes model-based error rate estimation techniques for HMM-based speech recognition systems and shows their effectiveness in several speaker-independent task conditions. The proposed techniques are based on an HMM-based error rate estimation method [C.-S. Huang *et al.*, IEEE Trans. SAP, **11**, No. 6, 581–589 (2003)], which only depends on subword-based pronunciation lexicon knowledge. This study assumes the need of the dynamic change of open vocabulary. The vocabulary size is varied between 5000 and 25 000 words. Two types of formalizations are investigated. They utilize the Bhattacharyya distance as a local distance measure between two competing word classes which are modeled by syllable-unit HMMs. A misclassification measure, which is similar to the one used by minimum classification error (MCE) training methods, is used for estimating error rate. The correlation between the error rates for testing utterances by 20 speakers and the estimated error rates for varied vocabulary sizes are evaluated. Except for one speaker, the correlation coefficient above 0.8 was obtained for both proposed formulations. Moreover, when the uttered words are clustered to five coarse classes of different estimated error rates, the correct rate of relative consistency between those adjacent classes was above 90% on average.

1pSC31. Large-scale continuous speech recognition system design using discriminative training. Erik McDermott and Atsushi Nakamura (NTT Commun. Sci. Labs, NTT Corp., Hikari-dai 2-4, Seikacho, Soraku-gun, Kyoto-fu 619-0237, Japan, mcd@cslab.kecl.ntt.co.jp)

Discriminative training is difficult to implement but essential to attaining state-of-the-art performance in automatic speech recognition systems today. Most of the discriminative training results for large scale recognition tasks (with vocabularies well over 10 000 words) so far use the maximum mutual information (MMI) framework, but recent results for the minimum classification error (MCE) framework suggest that MCE too yields significant improvements in recognition accuracy and system compactness on large-scale tasks. MCE embodies rather well the general intuition that recognition system design should attempt to improve performance (i.e., recognition accuracy) directly, by optimizing a criterion function that is closely related to performance, rather than indirectly, by optimizing a criterion such as overall log likelihood that does not reflect performance. This presentation provides an overview of the MCE framework, and describes recent MCE speech recognition results obtained with both the MIT Galaxy system and the NTT Communication Science Labs SOLON system. The tasks examined include a 33 000-word vocabulary telephone-based spontaneous speech weather information task, a 22 000-word telephone-based name recognition task, and a 100 000-word Japanese lecture speech transcription task.

1pSC32. Hyperparameter estimation for speech recognition based on variational Bayesian approach. Kei Hashimoto, Heiga Zen, Yoshihiko Nankaku, Lee Akinobu (Nagoya Inst. of Technol., Dept. of Comput. Sci. and Eng., Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan, bonanza@ics.nitech.ac.jp), and Keiichi Tokuda (Nagoya Inst. of Technol., Nagoya, 466-8555 Japan)

A hyperparameter estimation technique for HMM-based speech recognition using variational Bayesian (VB) approach is proposed. In recent speech recognition systems, the ML and MDL criteria have been adopted for training HMMs and the context clustering, respectively. However, since the ML estimator produces a point estimate of parameters, the estimation accuracy is reduced with insufficient data. The MDL criterion is also affected by the amount of training data due to an asymptotic approximation. To avoid this problem, the VB method has been applied to HMM-based speech recognition. By using the VB approach, it is expected that the higher generalization ability is achieved and an appropriate model structure can be selected in the context clustering. In the Bayesian framework, the predictive distribution is obtained by integrating the likelihood function with respect to the prior density. Therefore, the performance of the Bayesian method depends on the setting of hyperparameters representing the prior distribution. However, the detailed investigation for hyperparameter estimation has not been conducted in the speech recognition field. This study proposes an estimation technique of hyperparameters based on the maximization of the marginal likelihood for HMM-based speech recognition. Experimental results show that the proposed technique can estimate appropriate hyperparameters from training data.

1pSC33. An investigation of early reflection signal based on impulse response on automatic speech recognition. Yoshiki Hirano, Yuki Denda, Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, rs037021@se.ritsumei.ac.jp), and Takanobu Nishiura (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

Early reflection signals (especially, within 50 ms from direct sound arrival) are useful to speech listening for human being. On the other hand, the higher the reflection energy becomes, the more present automatic speech recognition performance decreases, subject to using clean acoustic phoneme model. However, it was not clear that early reflections are useful or useless to recognize the speech on distant-talking speech recognition, because reverberation time was employed as reverberation criterion. It is a unique time and does not depend on source position in a room. In addition,

it can only express the time in which sound attenuation becomes -60 dB and cannot express the early reflection that strongly influences speech recognition performance. Therefore, it should be nonsuitable reverberation criterion for distant-talking speech recognition. To overcome this problem, we investigate the relation between early reflections and distant-talking speech recognition performance based on impulse response between talker and microphone towards newly suitable reverberation criterion on distant-talking speech recognition. As a result of evaluation experiments in real acoustic environments, we could confirm the relation between speech recognition performance and early reflection signals. In future work, we will try to propose the newly suitable reverberation criterion. [Work supported by MEXT of Japan.]

1pSC34. Acoustic modeling with contextual additive structure for hidden Markov model-based speech recognition. Kazuhiro Nakamura, Heiga Zen, Yoshihiro Nankaku, and Keiichi Tokuda (Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan)

This study proposes an acoustic modeling technique based on an additive structure of context dependencies for HMM-based speech recognition. Context dependent models, e.g., triphone HMMs, are widely employed in speech recognition systems. Generally the dependency between phonetic contexts and acoustic feature distributions is direct, i.e., if a context is given, the Gaussian distribution is specified immediately. However, more complex dependencies can be considered. This study assumes an additive structure of acoustic feature components that have different context dependencies. For example, if the left and right phones are assumed as contexts of additive components, the generative process of the triphone feature is as follows: the additive components of left and right contexts are generated independently from each distribution and then added to the monophone feature. Since the output probability distribution is composed of the sum of the mean vectors and covariance matrices of additive components, a number of different distributions can be efficiently represented by a combination of fewer distributions. However, it is unknown which kinds of contexts have additive dependencies to acoustic features. Therefore, this study proposes a context clustering algorithm for the additive structure model. The proposed algorithm automatically extracts additive components by constructing multiple decision trees simultaneously.

1pSC35. Towards an efficient archive of spontaneous speech: Design of computer-assisted speech transcription system. Hiroaki Nanjo (Faculty of Sci. and Technol., Ryukoku Univ., Seta, Otsu 520-2194, Japan) and Tatsuya Kawahara (Kyoto Univ., Sakyo-ku, Kyoto 606-8501, Japan)

Computer-assisted speech transcription (CAST) system for making archives such as meeting minutes and lecture notes is addressed. For such a system, automatic speech recognition (ASR) of spontaneous speech is promising, but ASR results of spontaneous speech contain a lot of errors. Moreover, the ASR errors are essentially inevitable. Therefore, it is significant to design a good interface with which users can correct errors easily in order to take advantages of ASR for making speech archives. It is From these points of view that our CAST system is designed. Specifically, the system has three correction interfaces: (1) pointing device for selection from competitive candidates, (2) microphone for respeaking, and (3) keyboard. One of the most significant correction methods is selection from competitive candidates, thus more accurate competitive candidates are required. Therefore, generation methods of competitive candidates are discussed. Then, a speech recognition strategy (decoding method) based on the minimum Bayes-risk (MBR) framework is discussed. Since MBR decoding can distinguish and deal with content words and functional words, which are conventionally treated in a same manner, the approach is expected to generate more content words in competitive candidates.

1pSC36. Acoustic modeling of spontaneous speech of Japanese preschool children. Izumi Shindo, Tobias Cincarek, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma-shi, Nara 630-0192, Japan)

In recent years, there is an increasing demand for speech recognition of children. However, the recognition of children's speech, especially preschool children (2 to 5 years of age), is very difficult. For example, recognition accuracy using a children's acoustic model provided by the Japanese Dictation Toolkit is only 21.4%. Many different variations of child speech with palatal sounds and pronunciation error decrease recognition performance. This paper proposes a recognition method that investigates the characteristics of preschool children's speech using experimental data and considers phonetic changes. Mapping between standard and altered pronunciations of words is determined. In experiments, a large amount of spontaneous child speech (2 to 15 years of age) was collected with the speech-oriented public guidance system, "Takemaru-kun," which is currently available. Recognition performance increases to 49.2% by acoustic model adaptation of preschool children's speech. When allowing multiple pronunciation variations per word during recognition, further improvement to 52.0% is achieved.

1pSC37. Robust speech recognition for the control of wheelchairs by inarticulate speech of the severely disabled. Hiroaki Kojima, Akira Sasou, Soo-Young Suk, Shi-Wook Lee, and Ken Sadohara (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), Central 2, Umezono, Tsukuba, Ibaraki 305-8568, Japan, h.kojima@aist.go.jp)

The goal of this research is to develop voice-controlled wheelchairs that can be operated by inarticulate speech affected by severe cerebral palsy or quadriplegia, for instance. In this case, principal factors obstructing recognition performance are significant pronunciation variation caused by difficulty in stable articulation and bad influences of variety of noise in the real environment. To cope with the pronunciation variations, pronunciation lexicons that consist of multitemplates of reference patterns are utilized. The reference patterns are represented with subphonetic codes, which can describe variations of inarticulate speech more precisely than ordinary phonetic transcriptions. Pronunciation lexicons are generated by generalizing coded samples into compact representation of templates based on DP and data mining. For noise robustness, a voice activity detection method is investigated in order to circumvent friction of microphone, cough, etc. A sound source localization using a microphone array is also integrated in order to reject sounds from outside of the wheelchair. These methods are integrated into a system that can be mounted on a wheelchair. A usability test operated by a severe cerebral palsy patient in the real environment results in 95.8% accuracy within 1404 samples for the recognition of the five-command set. [Work supported by MEXT.]

1pSC38. Selection of recognition candidates based on parallel recognition of sentences and partial words in spoken dialogue. Makiho Kawakami (Faculty of Eng., Chiba Univ.; 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan), Hiromasa Terashi, Masafumi Nishida, Yasuo Horiuchi, and Akira Ichikawa (Chiba Univ., Inage-ku, Chiba 263-8522, Japan)

We have proposed a method that predicts a user's utterances in spoken dialogue systems by recognizing prediction sentences of each dialogue state, thereby decreasing recognition errors. However, the conventional method might repeat recognition errors because it confirms the whole sentence even if it recognizes only a part of a long word, such as a compound word, correctly. For this study, we introduce a method using decoders based on prediction sentences and partial words obtained by dividing long words. The proposed method can confirm a user's utterances by selecting a candidate to a partial matched word using recognition results of the prediction sentence and a divided partial word. We conducted experiments for 560 utterances of seven persons using a navigation system. Results showed that the word recognition accuracy was 81.1%. The rate at which recognition results by the prediction sentence and a divided

partial word were an exact or partial match was 70.5%. The rate of recognition errors when the recognition results were an exact or partial match was 0.9%. Therefore, we demonstrated that it is possible to confirm a user's utterances using the proposed method by selecting recognition candidates.

1pSC39. Multiple utterance prediction based on a tree structure of dialogue states. Hiromasa Terashi, Masafumi Nishida, Yasuo Horiuchi, and Akira Ichikawa (Grad. School of Sci. and Technol., Chiba Univ., 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan)

A problem of spoken dialogue systems is that recognition errors increase when the recognition-word vocabulary is large. To solve the problem, we propose a method to predict a user's utterances by recognizing prediction sentences of each dialogue state. The method can judge whether the user's utterance is within the range of prediction using decoders based on a prediction sentence and a large word vocabulary. However, the conventional method performs processing to lead in a prediction frequently when a topic changes because even an utterance within a domain is treated as an utterance of the prediction outside. In this study, we propose a new method using multiple predictions by performing recognition based on all prediction sentences in a domain. The proposed method produces a tree structure for dialogue states in a domain and regards child nodes of an arbitrary node as within the range of prediction. We conducted experiments for 451 utterances of 15 dialogues by ten persons using the university guide system. As a result, recognition accuracy of prediction sentences was 97.7% and the judgment accuracy of utterance predictions was 88.2%; high performance was obtained using the proposed method. Results demonstrate that it is possible to perform flexible dialogue control.

1pSC40. Document-level optimization in speech recognition. Rie Nakazato, Kugatsu Sadamitsu, and Mikio Yamamoto (Univ. of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki-ken, 305-8573, Japan)

Most ASR systems optimize scores at the sentence level because language models are designed to assign a meaningful probability to a sentence, but not to a document. However, in the last half-decade, NIPS people have specifically examined aspect-based global language models and developed generative text models such as the latent Dirichlet allocation (LDA), which can assign a meaningful probability to an entire document. We investigated a system to recognize all sentences in a read document, maximizing a global score including the document probability, in addition to assigning acoustic and local language probabilities. Document level optimization using global scores engenders a combinatorial problem in its implementation. Therefore, we used the rescoring framework with N-best recognition results for each sentence in a read document and the hill-climbing method to search for approximately optimal recognition results. In this study, we compare mixture of unigrams (MU), Dirichlet mixtures (DM), and LDAs as generative text models in experiments using read document speech data in Japanese Newspaper Article Sentences (JNAS) and academic presentation speech data in The Corpus of Spontaneous Japanese (CSJ).

1pSC41. Generating search query in unsupervised language model adaptaion using WWW. Yasutomo Kajiura, Motoyuki Suzuki, Akinori Ito, and Shozo Makino (Grad. School of Eng., Tohoku Univ., Aramaki, Aoba-ku, Sendai-shi, Miyagi, 980-9579 Japan, makino@ecei.tohoku.ac.jp)

To improve the accuracy of an LVCSR (large vocabulary continuous speech recognition) system, it is effective to gather text data related to the topic of the input speech and adapt the language model using the text data. Several systems have been developed that gather text data from World Wide Web using keywords specified by a user. Those systems require the user to be involved in the transcription process. However, it is desirable to automate the entire process. To automate the text collection, we propose a method to create an adapted language model by collecting topic-related

text from World Wide Web. The proposed method composes the search query from the first recognition result, and it gathers text data from the WWW and adapts the language model. Then the input speech is decoded again using the adapted language model. As the first recognition result contains recognition errors, we developed a method to exclude the misrecognized words using word-based confidence score and similarities between keywords. To evaluate the proposed method, we carried out adaptation experiments.

1pSC42. Collection of multimodal data in real-world driving. Takashi Kusakawa, Chiyomi Miyajima (Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan, kusakawa@sp.m.is.nagoya-u.ac.jp), Takanori Nishino (Nagoya Univ., 464-8603, Japan), Katsunobu Itou (Hosei Univ.), and Kazuya Takeda (Nagoya Univ., 464-8603, Japan)

Our research group has recently developed a new data collection vehicle equipped with various sensors for the synchronous recording of multimodal data including speech, video, driving behavior, and physiological signals. Driver speech is recorded with 12 microphones distributed throughout the vehicle. Face images and a view of the road ahead are captured with three CCD cameras. Driving behavior signals including gas and brake pedal pressures, steering angles, vehicle velocities, and following distances are recorded. Physiological sensors are mounted to measure the drivers' heart rate, skin conductance, and emotion-based sweating on the palm of the hand and sole of the foot. The multimodal data are collected while driving on city roads and expressways during four different tasks: reading random four-character alphanumeric strings, reading words on billboards and signs seen while driving, interacting with a spoken dialogue system to retrieve and play music, and talking on a cell phone with a human navigator using a hands-free device. Data collection is currently underway. The multimodal database will be published in the future for various research purposes such as noise-robust speech recognition in car environments, detection of driver stress while driving, and the prediction of driving behaviors for improving intelligent transportation systems.

1pSC43. Comparison of visual features for audio-visual speech recognition using the AURORA-2J-AV database. Takahiro Togo, Yukitaka Nimura, Takayuki Kitasaka, Kensaku Mori, Yasuhito Suenaga, Chiyomi Miyajima, and Kazuya Takeda (Dept. of Media Sci., Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan, ttogo@suenagam.is.nagoya-u.ac.jp)

Various techniques for noise-robust speech recognition have been developed for many years. However, speech recognition in nonstationary noisy environments, such as in car conditions, is still a very challenging research problem. Audio-visual speech recognition is a promising technique for improving noise robustness in such adverse acoustic conditions. Our research group has recently developed an audio-visual speech database "AURORA-2J-AV" following the AURORA2 database protocol, which has been widely used for evaluating the performance of audio-only speech recognition techniques. The AURORA-2J-AV database consists of audio speech signals and two types of image sequences: color and infrared human face images recorded when multi-digit numbers are spoken by about 100 native Japanese speakers. Visual noise as well as acoustic noise is added to the database after recording for simulating various conditions. Audio-visual speech recognition experiments are conducted using the audio speech and infrared images of the AURORA-2J-AV database. Two visual features, correlation coefficients (CCs) between two successive image frames and principal component scores (PCSs) of lip images, are compared in various signal-to-noise ratio (SNR) conditions. Experimental results show that CCs outperform PCSs at high SNRs, while PCSs function better at lower SNRs.

1pSC44. A method for estimating the degree of a speaker's anger using acoustic features and linguistic representation. Yoshiko Arimoto (Grad. School of Bionics, Comput. & Media Sci., Tokyo Univ. of Technol., 1401-1 Katakura, Hachioji, Tokyo, 192-0982, Japan), Sumio Ohno, and Hitoshi Iida (Tokyo Univ. of Tech., Hachioji, Tokyo, 192-0982, Japan)

Automatic speech recognition (ASR) systems are greatly demanded for customer service systems. With advanced interactive voice response systems, humans have more opportunities to have dialogues with computers. Existing dialogue systems process linguistic information, but do not process paralinguistic information. Therefore, computers are able to obtain less information during a human-computer dialogue than a human can during a human-human dialogue. This report describes a study of the estimation method of degree of speakers' anger emotion using acoustic features and linguistic representation expressed in utterances during a natural dialogue. To record utterances expressing the users' internal anger emotion, we set pseudo-dialogues to induce irritation arising from discontentment with the ASR system performance and to induce exasperation against the operator while the user makes a complaint. A five-scale subjective evaluation was conducted to mark each utterance with a score as the actual measurement of anger emotion. As a result of this, an emotional speech corpus was produced. We examine the acoustic features and features of linguistic representation of each utterance with reference to these anger score. Then we conduct experiments to estimate automatically the degree of anger emotion using parameters selected from those features.

1pSC45. Marking up Japanese Map Task Dialogue Corpus. Tomoko Ohsuga, Shuichi Itahashi (Natl. Inst. of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan, osuga@nii.ac.jp), and Syun Tutiya (Chiba Univ.)

This presentation reports an outline of marking up the Japanese Map Task Dialogue Corpus. The project was conducted by an independent group of researchers from the faculties of Chiba University and other institutions. This corpus contains 128 dialogues by 64 Japanese native speakers. The basic design of the dialogues and recordings conform to the original HCRC Map Task Corpus of the University of Edinburgh. Two speakers participated in the map task dialogues; an instruction giver who has a map with a route and an instruction follower who has a map without a route. The giver verbally instructs the follower to draw the route on his/her map. The corpus is marked up according to XML-based TEI (Text Encoding Initiative) P5 format, which is developed in an effort to provide interchangeable and shareable text data. The transcriptions of TEI format are viewed as "tags" that describe the start and end times of utterances, the duration of pauses, nonverbal events, and synchronization of overlapping utterances. This corpus will be available through NII-SRC (Speech Resources Consortium).

1pSC46. Towards the detection of potentially hazardous situations in vehicle traffic using driver speech and brake pedal operation. Lucas Malta (Grad. School of Information Sci., Nagoya Univ., Nagoya-shi, Chikusa-ku Higashiyama-doori 5-10 Nakanishi biru 306 - Japan, malta@sp.m.is.nagoya-u.ac.jp), Chiyomi Miyajima (Nagoya Univ., Japan), Katsunobu Itou (Hosei Univ., Japan), and Kazuya Takeda (Nagoya Univ., Japan)

A method for automatic detection of potentially dangerous situations in motor vehicle traffic is introduced. Unlike precedent works, which typically relied on camera arrays or road-traffic monitoring sensors to detect collision incidents, the proposed approach specifically incorporates changes in a drivers' behavior, detected through driver speech and brake pedal operation. Experiments were performed using a large real-world multimedia driving database of 493 drivers, obtained from the Centre for Integrated Acoustic Information Research (CIAIR, Nagoya University). The drivers, who interacted verbally with a human operator, uttered expletive words to express negative feelings in 11 of the 25 situations that we selected as potentially hazardous. In 17 of them, sudden and intense com-

pression of the brake pedal was observed. The proposed lexicographical speech-feature-based method also detected 33 false alarms to detect 80% of these 11 scenes. As for the other 17 scenes, our method based on two-dimensional histograms of brake pressure and its dynamics achieved an 80% detection rate for 473 false alarms. Analyses of data recorded while drivers interacted with a machine and a Wizard of Oz system as well as a rank of the most commonly uttered words in dangerous situations are also presented.

1pSC47. Computational complexity of a distance-based active search algorithm. Masahide Sugiyama (The Univ. of AIZU, Tsuruga, Ikki-machi, Aizu-Wakamatsu-shi, Fukushima, 965-8580 Japan)

An efficient pattern (segment) search in a large-scale database is a very important technology in current widespreaded Internet society. To achieve efficient searches, many efficient pattern search algorithms have been proposed. Active search (AS) is one of them; it is based on skipping of search calculation. Initially, AS was formulated between two histograms using a similarity measure. In fact, AS can be formulated using L_p distance, where p is greater than 1. The search efficiency can be evaluated using a number of distance calculations. The first result in this paper shows that the L_1 distance in AS gives the minimum number of distance calculations among L_p distances. This property can be derived from a newly derived inequality between the L_p and L_1 norm on output probability space. A new distance can be derived from positive weighting of any distance. The second result of this paper is that the average skip width for any weighed L_p distance cannot be greater than the most efficient distance. Therefore, no positive combination of L_p can be more efficient than L_1. The two results indicate that L_1 distance is the most efficient for active searching.

TUESDAY AFTERNOON, 28 NOVEMBER 2006   HONOLULU ROOM, 1:00 TO 4:50 P.M.

Session 1pSP

Signal Processing in Acoustics: Blind Signal Processing

Leon H. Sibul, Cochair

Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Shoji Makino, Cochair

NTT Communication Science Lab., 2-4 Hikaridau, Seika-cho, Sorakugun, Kyoto, 619-0237, Japan

Chair's Introduction—1:00

Invited Papers

1:05

1pSP1. Frequency domain blind source separation in a noisy environment. Ryo Mukai, Hiroshi Sawada, Shoko Araki, and Shoji Makino (NTT Commun. Sci. Labs., NTT Corp., 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan)

A prototype system for blind source separation (BSS) in a reverberant and noisy environment is presented. Our system uses a small three-dimensional array with eight microphones and realizes BSS in various situations such as the separation of many speech signals located in a three-dimensional space, and the extraction of primary sound sources surrounded by many background interferences. The mixed signals observed by the microphone array are processed by independent component analysis (ICA) in the frequency domain. The permutation problem is solved by normalized basis vector clustering, which is a generalized version of directions of arrival (DOA) clustering and has several advantages to the DOA based method. The system estimates the DOA of the source signals as a by-product of the separation process. Moreover, the system has the ability to distinguish primary target signals and ambient noise. Our system performs an on-the-spot BSS of live recorded signals. Live demonstration will be performed at the meeting.

1:25

1pSP2. Beyond the independent component analysis: New blind acoustic sound separation in real world via single-input multiple-output-based independent component analysis. Hiroshi Saruwatari (Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma, Nara, 630-0192 Japan)

This review discusses a new generation of blind acoustic-signal separation in the real world, focusing on a recently proposed ICA method, Single-Input Multiple-Output (SIMO)-signal-based ICA (SIMO-ICA). The term "SIMO" is referred to as a specific transmission system in which the input is a single source signal and the outputs are its transmitted signals observed at multiple microphones. The SIMO-ICA consists of multiple ICA parts and a fidelity controller, and each ICA runs in parallel under the fidelity control of the entire separation system. In the SIMO-ICA scenario, mixed multiple sources detected at the microphones can be separated, not into monaural source signals but into SIMO-model-based contributions from independent sources as they are at the microphones. Thus, the separated signals of the SIMO-ICA can maintain the spatial qualities of each sound source. This attractive feature of the SIMO-ICA shows the promise of applicability and connectability with the other multichannel signal processing for achieving the higher performance beyond the conventional ICA methods. As the successful examples of SIMO-ICA's application, three kinds of systems are described: (*i*) combination with inverse filtering for MIMO decomvolution, *ii*) combination with adaptive beamformer for higher-performance BSS, and *iii* combination with binary masking for real-time processing.

1pSP3. How blind are we in blind signal processing? Walter Kellermann (Univ. Erlangen-Nuremberg, Cauer-str. 7, 91058 Erlangen, Germany, wk@lnt.de)

Blind signal processing algorithms aim at blindly solving classical signal processing problems such as system identification, signal separation, or source localization. Here it is investigated to which extent some of the more popular blind signal processing concepts are really blind when they are used for signal acquisition in the acoustic domain. Blindness regarding the source configuration and the source signal properties, blindness with respect to channel properties, and, finally, blindness regarding the microphone configuration are investigated. It turns out that many algorithms are explicitly nonblind with respect to important properties of the source configuration, such as the number of sources and their pointlike nature. The source signals typically have to fulfill certain statistical properties or have to meet sparseness constraints. Some algorithms rely on a dominant direct acoustic path between sources and sensors, and linearity of the transmission channel model is implied with all convolutive mixture models. Finally, some popular algorithms are actually aiming at direction of arrival estimation, which always requires knowledge on the microphone array geometry. In essence, blind algorithms are wellinformed in many respects and are blind only with regard to a few, although decisive, properties.

1pSP4. Informed acoustic source separation and localization. Kevin H. Knuth (Dept. of Phys., Univ. at Albany, Albany, NY 12222)

Advances in Bayesian computational technology in the last decade have enabled the development of new source separation and source localization algorithms. These algorithms are greatly improved by the encoding of prior information about a specific problem in the form of the chosen relevant model parameters, the assignment of the likelihood functions, and the assignment of the prior probabilities of the model parameter values. I refer to such source separation algorithms as informed source separation for the reason that they are endowed with specific and often vital information about the problem. Furthermore, the Bayesian methodology allows source separation to be united with source localization simply by including the model parameters that are of interest to the researcher. Here, I will discuss the union of source separation and source localization under the Bayesian methodology, the incorporation of prior information, and the construction of an informed algorithm using the new computational technologies that allow us to estimate the values of the parameters that define these high-dimensional problems.

1pSP5. Recovering the quality of speech degraded by reverberations in a room. Masato Miyoshi (NTT Commun. Sci. Labs, 2-4 Hikaridai Seika-cho, Keihanna Sci. City, Kyoto 6190237 Japan, miyo@cslab.kecl.ntt.co.jp)

Speech signals captured by distant microphones in a room are usually reverberated due to wall reflections. Reverberation may seriously deteriorate the signal characteristics, thus damaging the quality of such applications as hands-free telephony or automatic speech recognition. To eliminate such room reverberation effects, the inverse-filtering of the room transfer functions (RTFs) between a speaker and the microphones appears to be a very promising approach. When applied to a microphone system, this processing should work blindly because no explicit information on RTFs or source signals is available. The blind design of an inverse-filter set may be roughly classified into three approaches: (1) calculating an inverse-filter set of RTF estimates obtained from observed reverberant signals based on the ''subspace method,'' (2) calculating the filter set from observed signals and replicas of direct-sound signals using statistical properties inherent in speech signals, and (3) calculating the filter set directly from received signals, by whitening these signals without excessive degradation of the original signal characteristics. This presentation will provide an overview of these three approaches and introduce our speech dereverberation trials.

1pSP6. Blind deconvolution for noisy dynamic channels. Michael J. Roan (Dept. of Mech. Eng., Virginia Tech., Blacksburg, VA 24060)

It is common in acoustics to measure a signal that has been degraded by propagation through an unknown, noisy channel prior to measurement. While only the degraded measured signal is available, the data of interest are the original signal and the channel parameters. Often, it is desirable to reverse the filtering process by application of an inverse filter to recover the original signal. When neither the input signal properties nor the channel properties are known, this process is known as blind deconvolution (BDC). Typically, BDC algorithms assume noiseless, stationary propagation channels and input sources. These assumptions are usually violated in practical applications (e.g., noisy multipath propagation environments with moving source and receiver). To model these effects, predictive techniques are applied to incorporate *a priori* information about the system into the existing blind processing framework. The original contributions of this work follow. First, a novel formulation of the extended Kalman filter (EKF) is proposed. This allows incorporation of *a priori* information into gradient-based blind processing algorithms. This formulation is then applied to the natural gradient (NG) BDC algorithm. Finally, results are presented that suggest significant improvement in signal recovery performance over the NG BDC algorithm for dynamic noisy channels.

Contributed Papers

3:20

1pSP7. Blind dereverberation based on auto-correlation functions of framewise time sequences of frequency components. Kenko Ota and Masuzo Yanagida (Doshisha Univ., 1-3, Tarata-Miyakodani, Kyotanabe, Kyoto, 610-0321, Japan, etf1704@mail4.doshisha.ac.jp)

Proposed is a new blind dereverberation method based on auto-correlation function of framewise time series of each frequency component. Inverse filtering of source-microphone transfer functions is widely employed for suppressing the effects of reflected waves, but this method cannot be employed for cases where source-microphone transfer functions are not available or for time-variant cases. Several methods of spectral subtraction have been proposed to cope with these cases. However, most of them require transfer functions among sources and microphones. The proposed method, however, does not require transfer function inverse filters. Moreover, the method can handle frequency characteristics of sound paths. To realize dereverberation, it is necessary to estimate the delay time and frequency characteristics of reflection, though most conventional methods assume flat frequency characteristics. The proposed method estimates these parameters based on auto-correlation function of each time series of spectral components in frequency spectra obtained every frameshift interval. The proposed method improves dereverberation performance, comparing with methods that assume flat frequency characteristics. The proposed method raises average segmental SNR by 3.4 dB and reduces reverberation time from 390 to 109 ms for a room employed for experiments. [Work supported by Knowledge Cluster Project, MEXT, and by Academic Frontier Project Doshisha University.]

3:35

1pSP8. Sparseness-based 2ch blind source separation in reverberant environments based on direction-of-arrival estimation with a reliability measure. Yosuke Izumi, Nobutaka Ono, and Shigeki Sagayama (Dept. of Information Sci. and Technol., The Univ. of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, izumi@hil.t.u-tokyo.ac.jp)

We present a novel sparseness-based 2ch blind source separation (BSS) method based on robust direction-of-arrival (DOA) estimation that enables reasonable separation of source signals even in a highly reverberant environment. DOA estimation for BSS is not easy in reverberant environments since reflected sound waves behave as coherent interferences in the observed signals and make ambiguous the sound source direction. Our key idea to solve this problem is to find and integrate clean, reliable time-frequency fragments through the observation with a reliability measure. For that, (1) based on a diffused sound field model of reverberation, a theoretical correlation matrix is introduced for whitening the coherent interferences, (2) the DOA of each time-frequency bin is estimated with reliability by utilizing information of neighboring bins, and (3) the reliability-weighted DOAs are integrated over the whole time-frequency bins. After that, separated signals are obtained by time-frequency masking or inverse filtering. Experimental results including comparison with other methods are also reported.

3:50

1pSP9. Multilayered spatio-temporal gradient analysis for acoustic blind source separation. Kenbu Teramoto, Md. Tawhidul Islam Khan, Seiichirou Torisu, and Akito Uekihara (Saga Univ., 1-Honjo, Saga-shi, Japan, 8408502, tera@me.saga-u.ac.jp)

A novel blind source separation of a mixture of two or more voice signals has been proposed in the present paper. The separation system has been focused based on the spatio-temporal gradient analysis. The proposed algorithm utilizes the linearity among the signals: sound pressure of source signals, the three-dimensional (x, y, and z directional) particle velocity vector, and its gradient of the observed signals, all of which are governed by the equation of motion. Principally, as the mechanism of blind source separation uses no-*priori* information about the parameters of convolution, filtering as well as mixing of source signals, some simple assumptions such as the statistical independency of the linearly combined (mixed) observed signals containing zero mean as well as unit variance have been implied in the present separation algorithm. Therefore, the proposed method has successfully simplified the convoluted blind source separation problem into an instantaneous blind source separation problem over the spatio-temporal gradient spaces. An acoustic experiment with two female voices has been carried out to compare the simulated data as well. A eight-microphone array system has been adopted to evaluate the voice signals efficiently.

4:05

1pSP10. Study for blind source separation on time-frequency domain considering phase information. Fumio Sasaki, Masahito Yasuoka, and Osamu Tanaka (Dept. of Architecture, Tokyo Univ. of Sci., 1-3 Kagurazaka, Shinjuku-ku, Tokyo 1628601, Japan, fsasaki@rs.kagu.tus.ac.jp)

The methods that can be done for the estimation of the number of source signals and separation of source signals using time-frequency information are proposed as a method of blind source separation. In these methods, some independent conditions are assumed in a time-frequency domain, and observed signals are expanded to time-frequency domain using wavelet analysis. However, these methods are not considered phase differences (difference of distance) between source signals and observed signals. A new method is proposed. The method can be done not only for the source separation but also the specification of the locations of source signals considering phase differences. In this method, it is necessary to assume a rather stronger independent condition than former methods. But, on account of this assumption, a function that becomes a real value only when phase differences are coincidence in time-frequency domain is determined. Using this function, first of all, the number of source signals is calculated, then the locations of source signals and source signals are calculated. The effectiveness of the method is shown using actual synthetic signals.

4:20

1pSP11. Blind spatial subtraction array based on independent component analysis for speech enhancement and recognition. Yu Takahashi, Tomoya Takatani, Hiroshi Saruwatari, and Kiyohiro Shikano (Speech and Acoust. Processing Lab., Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192 Japan)

We propose a new blind spatial subtraction array (BSSA) that contains an accurate noise estimator based on independent component analysis (ICA) for the realization of noise-robust hands-free speech recognition. Many previous studies on ICA-based blind source separation often dealt with the special case of speech-speech mixing. However, such a sound mixing is not realistic under common acoustic conditions; the target speech can be approximated to a point source but real noises are often not point sources. Under the condition, our preliminary experiment suggests that the conventional ICA is proficient in the noise estimation rather than the direct speech estimation. Based on the above-mentioned findings, we propose a new noise reduction method that is implemented in subtracting the power spectrum of the estimated noise by ICA from the power spectrum of noise-contaminated observations. This architecture provides us a noise-estimation-error robust speech enhancement rather than a simple linear-filtering-based enhancement. Although nonlinear processing often generates an artificial distortion, the so-called musical noise, it is still applicable to the speech recognition system because the speech decoder is not so sensitive to such a distortion. Experimental results reveal that the proposed BSSA can improve the speech recognition rate by 20% compared with the conventional ICA.

1pSP12. Frequency-domain independent component analysis by overlap piecewise integration of separation processing. Tadashige Noguchi, Kenko Ota, Masuzo Yanagida (Doshisha Univ., 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto, 610-0321 Japan, dtg0731@mail4.doshisha.ac.jp), and Leandro Di Persia (Universidad Nacional de Entre Ros Casa de la Universidad, Paran, Entre Ros, Argentina)

Conventional frequency-domain ICA yields the optimal separation for each frequency bin, but it suffers from the permutation problem. The authors have developed permutation-free ICA as a separation scheme by obtaining the separation matrix for a long vector consisting of temporal changes of all frequency components of the received signal. The permutation-free ICA, however, only yields a common separation matrix for all frequency bins. So, the separation matrix obtained in the permutation-free ICA has a common directivity pattern for all frequency bins, though the method can avoid the permutation problem. Proposed in this paper is a scheme of multibin ICA that deconvolves mixed signals into original source signals by shifting piecewise integration of a set of frequency bins consisting of the frequency bin in concern and neighboring frequency bins. The proposed method can yield nearly optimal directivity in the form of the separation matrix for each frequency bin, avoiding the permutation problem. Performance of multibin ICA is compared with those of conventional frequency-domain ICA and permutation-free ICA employing segmental SNR as an evaluation index. [Work supported by Knowledge Cluster Project, MEXT, and by Academic Frontier Project Doshisha University.]

TUESDAY AFTERNOON, 28 NOVEMBER 2006　　　　　WAIALUA ROOM, 1:00 TO 5:35 P.M.

Session 1pUW

Underwater Acoustics: Array Processing, Sensors, and Technology

Paul Hursky, Cochair
Heat Light and Sound Research, 12730 High Bluff Dr., San Diego, CA 92130

Hiroyuki Hachiya, Cochair
Chiba Univ., Research Ctr. for Frontier Medical Engineering, 1-33 Yayoi-cho, Inake-ku, Chiba, 263-8522, Japan

Chair's Introduction—1:00

Contributed Papers

1:05

1pUW1. Acoustic seaglider. Bruce M. Howe (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, howe@apl.washington.edu)

Continually improving acoustics measurements in the ocean is a *sine qua non* for the advancement of ocean acoustics and related fields. Acoustic sensors on mobile platforms will enable basic research topics on temporal and spatial coherence and the description of ambient noise, with direct impact on the applicability of coherent processing with its associated gain to signal detection and acoustic navigation and communications within the context of distributed undersea sensor networks. We describe the integration of acoustic receiving and communication capability in gliders with results from several recent field tests. [Work supported by the ONR.]

1:20

1pUW2. Single element synthetic aperture using an ocean glider. Georges A. Dossot, James H. Miller, Kristy A. Moore (Dept. of Ocean Eng., Univ. of Rhode Island, 217 Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882, georges@oce.uri.edu), Steven E. Crocker (Naval Undersea Warfare Ctr., Newport, RI 02871), Jason D. Holmes (Boston Univ., Boston, MA 02215), and Scott Glenn (Rutgers Univeristy, NB, NJ 08901-8521)

The feasibility of using a single transducer element on an ocean glider to create a synthetic aperture is discussed. Acoustic data were collected on two Webb Slocum gliders deployed by Rutgers University during the Shallow Water Experiment (SW06) on the continental shelf off New Jersey. These gliders periodically surfaced for GPS fixes and data transfer via satellite phone. A synthetic aperture is created through coherent processing of the acoustic data as the glider travels through the water. A number of issues including changes in depth, nonsteady motion of the glider, and clock drift can limit the performance of the processing. However, the glider provides a low-noise and low-speed platform, potentially improving the signal-to-noise ratio. The geometry of the experiment provided for near-broadside recording of the 224- and 400-Hz WHOI tomography sources. An acoustic normal mode representation of the field provides the basis for processing of the data similar to the Hankel transform approach of Frisk and Lynch [J. Acoust. Soc. Am. **76**, 205–216 (1984)] and Holmes *et al.* [J. Acoust. Soc. Am. **119**, 3346 (2006)]. Effects of spatial variations in sound speed are examined. [Work sponsored by the Office of Naval Research.]

1:35

1pUW3. Processing data from a low-drag array towed by a glider. Paul Hursky, Michael B. Porter, Martin Siderius (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130), Vincent K. McDonald, Mark Gillcrist, Brian Granger, Ryan Jones, Aaron Bratten, Andy Huizinga, Peter T. Sullivan, and Susan G. Briest (Space and Naval Warfare Systems Ctr., San Diego, CA 92152)

Small underwater robotic vehicles such as gliders and AUVs have limited capabilities in terms of propulsion, speed, and power consumption, so it takes great care to design suitable sensor systems such as towed arrays for these platforms. Two such arrays were deployed during the Makai experiment, off the coast of Kauai in Hawaii. A 15-element line array with an acoustic aperture of 21 m was towed by a Webb Research Slocum glider. A 40-element line array with an acoustic aperture of 60 m was towed by a small work boat. We will describe the design and construction of these arrays and present results of processing data recorded on these arrays. We will also discuss potential applications for these arrays deployed from gliders and AUVs.

1pUW4. Synthetic stationary array processing for a towed array in maneuver. Yung Lee (Sci. Applications Intl. Corp., 1710 SAIC Dr., McLean VA 22102), William Lee (Duke Univ.), and Andrew Lee (MIT)

When long towed arrays undergo significant maneuvering, the array shape rapidly changes within a processing interval that degrades the array beamforming performance. Towed arrays are known to behave much like pulleys through the water. Thus, hydrophones on the towed array will closely follow the motion of the towed point through the water with a time lag relative to the towing speed. From this concept, it follows that, given an appropriate time alignment, a signal snapshot from the rear section of the array can be interpreted as a signal snapshot from the front section of the array at a later time. This synthetic stationary array processing (SSAP) divides the physical array into highly overlapping sequential segments and, using these segments, forms time-lagged snapshots of a stationary array. As a result, this synthetic stationary array has a reduced size and shape that is defined by the first array segment at the beginning of a processing interval. To demonstrate the effectiveness of this process, real towed array data collected from the MAPEX 2000 experiment conducted by SCALANTCEN were processed. The SSAP results successfully demonstrate processing improvements through turns. A few selected signal tracks were simulated to validate the performance of the SSAP processing.

1pUW5. Experimental results of a fast Fourier transform-based motion compensation of synthetic aperture sonar. Takao Sawa, Taro Aoki, Jyunichirou Tahara (Adv. Marine Technol. Program, JAMSTEC, 2-15 Natsuhima, Yokosuka, Kanagawa 2370061, Japan, sawa@jamstec.go.jp), Tomoo Kamakura, Ryouji Suzuki (Dept. of Elec. Eng., UEC), and Hitoshi Yamaguchi (Nagasaki Prefectural Inst. of Public Health and Environ. Sci.)

Synthetic aperture and its motion compensation technique became available in sonar systems owing to dramatic developments of computer performance. A bottom observation test with synthetic aperture sonar was carried out at the gulf of Katagami in November 2005. The sonar on a tow-fish had detected a bottom and floats at the test. Accelerations and angular velocities of the sonar system had been recorded with external sensors on the tow-fish, too. A relative position and attitude of sonar were calculated from recorded motion data, and a new motion compensation method based on FFT was applied to the synthetic aperture sonar algorithm. In this report, details and capacity of this new method are shown through these experimental results. Sonar images with this new compensation were better than a conventional compensation that uses full bandwidth of motion frequency. Moreover, images were improved by adjusting bandwidth. The authors would like to thank Takeshi Hamazone, a captain of the mother ship, for his maneuvering the ship even in bad weather.

1pUW6. Higher frequency matched-field processing. Claire Debever and William A. Kuperman (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093-0238, cdebever@ucsd.edu)

We study matched-field processing (MFP) in the 3.5-kHz shallow water regime at a range of 4 km for which MFP is typically problematic. Data are used to construct both the cross-spectral-density-matrix (CSDM) and the weight vectors for the Bartlett, minimum variance distortionless response (MV), and white noise constraint (WNC) processors. There is sufficient data to construct the weight vectors as a function of time as measured around the CSDM construction time. While the MV processor degrades significantly over a few minutes, the WNC processor remains stable for 15 min or longer. We study these experimentally derived weight vectors to further understand the origin of the increased stability of the WNC over the MV and to gain further insight into how to construct a robust MFP processor. [Work supported by ONR.]

1pUW7. Steering vector sensor array elements with cardioids, hippopedes, and other beam shapes. Kevin B. Smith (Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943) and A. Vincent Van Leijen (Netherlands Defense Acad., Den Helder, Netherlands)

The purpose of this report is to study the impact on linear beamforming due to individual vector sensor element steering patterns. Standard, linear beamformers employ cardioid beampatterns for each vector sensor. In this work, we examine a class of vector sensor element steering patterns beyond the standard cardioid. The element weighting is nonlinear but nonadaptive, making it simple to implement in hardware processing. The new sensor steering patterns, referred to as hippioids, are products of cardioids and various orders of hippopedes. The angular resolution of individual sensors, and the impact on angular resolution from arrays of varying aperture, will serve as the performance measure along with peak-to-sidelobe levels. An example of the differences in vector sensor steering patterns is provided using measured, DIFAR buoy data. [Work supported by the Office of Naval Research, Code 321OA, the Naval Postgraduate School, and the Netherlands Defense Academy.]

1pUW8. Wideband acoustic communication using quadrature-phase shift keying and eight-phase shift keying. Hiroshi Ochi, Yoshitaka Watanabe, and Takuya Shimura (Marine Technol. Ctr., JAMSTEC, 2-15 Natsushima-cho, Yokosuka, 237-0061 Japan, ochi@jamstec.go.jp)

In JAMSTEC, a short-range (up to 500 m) high-speed acoustic communication system for deep sea environment has been investigating. To widen a bandwidth, the new wideband projector was developed. It consists of two transducer elements and has a toroidal beam pattern. To reduce the acoustic radiation pattern of the rear side, the unit, which is like a lampshade, was attached to that projector. Moreover, the beam axis is tilted to the upper side by delayed input signal to the upper transducer element. When this projector is used in the vicinity of the sea bottom, a bottom-reflected wave can be reduced. Carrier frequency of this projector is 80 kHz, and its bandwidth is 40 kHz. The experiment was carried out at the depth of 1000-m area using this projector. Two omni-directional hydrophones were used for receiving. Quadrature phase shift keying (Q-PSK) and 8-PSK were used as its modulation method. The data of several transmission speeds [10 k symbol per second (sps), 12.5, 16, 20, 25, and 40 ksps] were obtained at the transmission range of 200–230 m. Under those conditions, error free communication was carried out for 60 000 bits of data.

1pUW9. The study on data transmission with short positioning pulse in deep sea. Yoshitaka Watanabe, Hiroshi Ochi, and Takuya Shimura (Japan Agency for Marine-Earth Sci. and Technol., 2-15, Natsushima-cho, Yokosuka-shi, Kanagawa pref, Japan, yoshitakaw@jamstec.go.jp)

The ultra short base line (USBL) underwater acoustic positioning in deep sea with digital data transmission is being studied. The goal of this study is improvement of the accuracy and expansion of the functions of the USBL for autonomous underwater vehicles (AUVs) and submersibles due to utilization of the transmitted data. To achieve the goal, the reliable data transmission with short positioning pulse must be ensured. The direct sequence spread spectrum (DSSS) is applied. The frequency band width of signal is 3 kHz with a center focus on 12 kHz for the deep sea use. From the ocean experimental data, it seems that the multipaths from the neighborhood have a dominant effect on the transmission quality. The results of investigation about this matter by experiments and simulations are shown. [This study is supported by a Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology in Japan.]

1p TUE. PM

3:35

1pUW10. Development of a digital 3-D SBL underwater positioning system combined with GNSS compass. Yoshinori Miyamoto, Masahiko Furusawa, Keiichi Uchida, Toshikazu Motoyoshi (Tokyo Univ. of Marine Sci. and Technol., miyamoto@s.kaiyodai.ac.jp), Toshihiro Yamaguchi, Katuji Miwa (Koden Electron. Co., Ltd.), and Toshiharu Kakihara (Tokyo Univ. of Marine Sci. and Technol.)

A highly digitized underwater positioning system using the SBL method was developed. This system can measure highly accurate three-dimensional positions of marine organisms equipped with an ultrasonic transmitter (pinger). This system is combined with a GNSS (GPS) compass that measures the bearing, the location (latitude and longitude), and the commotion (rolling angle and pitching angle) of a boat or a buoy. Then the pinger positions are compensated for the commotion and are translated into the underwater absolute positions. The four arbitrary frequencies of the pingers are received by four wideband hydrophones. The received signals are immediately AD converted in high speed and processed by the quadrature detection and the digital filtering with a bandwidth of 3 kHz using FPGA. Finally, the receiving time differences are measured by the cross correlation, realizing a resolution of 0.01 ms. A position measurement experiment was performed in a water tank (length 15 m, width 10 m, depth 10 m), changing the position of a pinger, and it is confirmed that almost the expected accuracy could be realized. Moreover, this system was installed in a boat and the absolute positions of a fish were pursued to confirm the practicality.

3:50

1pUW11. Continued Haro Strait path data processing. A. Tolstoy (ATolstoy Sci., 1538 Hampton Hill Circle, McLean, VA 22101) and Z.-H. Michalopoulou (NJIT, Newark, NJ)

Previous efforts have led to the estimation of candidate source-array (S-R) geometries for some Haro Strait data. In particular, nw014 (one broadband shot to the NW array) has been extensively analyzed with the result that source range, source depth, array phone localizations, water depth at the source (D1), and water depth at the array (D2) can now be restricted to sets based on the time-domain data alone, i.e., on the signal boundary reflections. Most recent efforts include FFTs of the data followed by MFP geoacoustic inversion at multiple high frequencies (480–520 Hz)—using the restricted sets—in order to estimate array depths and shapes as well as some of the bottom properties for the nw014 data. Array shape and phone localizations are extremely important but difficult to attain. This presentation will describe these new efforts.

4:05

1pUW12. Near-orthogonal signals. John Piper (NSWC-PC, 110 Vernon Ave., Panama City, FL 32407)

A least-squares approach was used to construct near-orthogonal signals that can be useful in sonar applications. For the matched-filter case the solution can be written with Lagrange multiplier terms, which reduces to a simple eigenvector problem. Decoding filters are also examined, and these results are compared to the matched-filter case.

4:20

1pUW13. Wave front design by acoustic contrast control. Joung-Woo Choi and Yang-Hann Kim (Ctr. for Noise and Vib. Control(NOVIC), 4114 ME Dept., KAIST, Sci. Town, Daejeon, Korea, jungwoochoi@gmail.com)

A way to design and generate desired wave fronts using multiple sound sources is proposed. In principle, any kind of wave front can be described as a sum of orthogonal basis functions, and the wave front of a specific shape can be generated by focusing multiple sources' energy on a single basis function. Once the sound field is decomposed into a set of orthogonal functions, we can apply a conventional focusing algorithm to generate a desired wave front. Extending the concept of energy focusing, we can also generate a wave front packet, whose energy is concentrated on a group of orthogonal functions rather than single component. This enables us to generate a wave front within a desired bound. For this purpose, we employ acoustic contrast control, which focuses sound energy on a selected group of orthogonal functions by enhancing the energy difference between the focal region and others. Various example cases demonstrate how the proposed method can generate a group of planar and spherical wave fronts propagating to different directions. [Work supported by the NRL project of KISTEP and BK21 project initiated by Ministry of Education and Human Resources Development of Korea.]

4:35

1pUW14. Predicted versus measured sensitivities of fiber-optic sensors. Fred C. DeMetz, Sr. (Sabeus, Inc., 26610 Agoura Rd., #100, Calabasas, CA 91302)

Fiber-optic sensors are finding wide use in underwater and seismic applications due to their ease of construction, wide useful acoustic bandwidth, sensor deployment ranges from the light source of 10 km or more, robust temperature and static pressure tolerance, and immunity from requiring supporting electronics near the measurement points. The theoretical sensitivity for a number of optical sensor designs are compared with experimental data. Their signal-to-noise performance is assessed from the standpoint of minimum detectable signal, environmental background, and optical noise source.

4:50

1pUW15. Fiber optic interferometric hydrophone based on fiber Bragg grating with polarization diversity receiver. Chiaki Okawara and Kenji Saijyou (5th Res. Ctr., Tech. R&D Inst., Japan Defense Agency, 3-13-1 Nagase, Yokosuka, Kanagawa, 239-0826 Japan)

The objective of this work is to overcome polarization-induced signal fading (PSF) that occurs in fiber optic interferometric hydrophones based on fiber Bragg grating (FBG). The use of FBGs as reflectors in fiber optic interferometric hydrophones would greatly simplify the construction of the devices' wet-end portion by eliminating couplers, joints, and other connections that are required for other types of fiber optic interferometric hydrophones. Unfortunately, the visibility of the interference signal is reduced by PSF, and the demodulated signal is contaminated greatly by the phase noise. A polarization diversity receiver (PDR) is therefore applied to reduce the PSF influence. We developed a prototype of fiber optic interferometric hydrophone based on FBG with the PDR and demodulate an acoustic signal to confirm the effectiveness of the PDR application. Experimental results described here show that the phase noise is suppressed by selecting an optimum port that has high visibility. Consequently, the effectiveness of the application of the PDR is confirmed.

5:05

1pUW16. A new optical-fiber-based underwater sensor. Francois M. Guillot, D. H. Trivett, and Peter H. Rogers (George W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

A new type of underwater sensor based on optical fiber Bragg gratings is presented. This transducer can be used in a dipole or a quadrupole (shear sensor) configuration. The sensing principle of the device relies on the interference signal from two Bragg gratings written on the same fiber and illuminated by a tunable, narrow-band light source. The gratings are 5 cm apart, and they each reflect a portion of the incident light. The fiber is epoxied to two spacers separated by a 1-mm gap situated between the gratings. This assembly is then adhered to two plates connected by a hinge, which is located below the gap. One plate is held rigidly and the tip of the other (free) plate experiences transverse vibrations, when ensonified. These vibrations produce periodic gap length changes, which modulate the interference signal from the two gratings. The modulation is related to the amplitude of the sound wave and is monitored with a photodetector. Design optimization to increase the sensitivity and to reduce the noise floor is discussed, and the performance of a practical sensor is presented. [Work supported by ONR.]

1pUW17. Development of an acoustic cloud condensation nuclei counter. Marcellin Zahui, Abhijit Deshpande, and David Delene (Univ. of North Dakota, Grand Forks, ND)

Theoretical development of an acoustic cloud condensation nuclei counter is presented. The proposed instrument will be able to count the number of aerosols present in a sample of air flowing through a growth chamber. The air condenses around aerosols and forms water droplets. The number of aerosols is determined by measuring the sound produced underwater by these droplets when the droplets strike a water surface at the bottom of the growth chamber with an impact velocity equal to either their terminal or maximum velocities. First, the terminal velocities of the droplets are calculated. Then, the maximum velocities that these droplets can sustain without breaking are calculated as a function of droplet diameter. Second, the sound due to droplet impact is estimated and experimentally verified. In this study the bubble sound is not considered because micro droplets falling with terminal velocities do not create bubbles. Also, when accelerated, the velocities are limited such that there is no bubble formation at impact. Finally, two approaches are considered for the instrument. One case uses one hydrophone to measure the total sound radiated by all the droplets. The other case is based on a grid of tiny hydrophones, each measuring a droplet impact sound.

1p TUE. PM

Session 2aAA

Architectural Acoustics: Acoustics of Large Indoor and Outdoor Performance Spaces I

Daniel R. Raichel, Cochair
2727 Moore Ln., Fort Collins, CO 80526

Toshiki Hanyu, Cochair
Nihon Univ., Narashinodai, Funabashi, Chiba 274-8501, Japan

Chair's Introduction—7:30

Invited Papers

7:35

2aAA1. Differences in dealing with the acoustics of large outdoor and indoor listening spaces. Daniel R. Raichel (Eilar Assoc. and the Graduate Ctr., CUNY, 2727 Moore Ln., Fort Collins, CO 80526, draichel@comcast.net)

It was perhaps fortuitous that Wallace Clement Sabine's first consulting assignment from the Harvard Board of Overseers, the execution of which proved to be seminal in the history of architectural acoustics, entailed an indoor facility. The problems of dealing with outdoor performance areas are considerably greater than those encountered in indoor spaces. Indoor spaces provide surface reflections and absorption of sound, both of which can be manipulated through proper design to yield the desired results with respect to reverberation time and sound dispersion. In the case of outdoor performing areas, there are no reflective surfaces (except in the cases of certain amphitheaters), and the need for sound amplification is often necessitated. A reflective bandshell is generally required to direct the program material toward the audience. The lack of wall reflections tends to result in a lack of musical sonority that is usually present in a well-designed indoor auditorium. A number of specific problems with interior and outdoor spaces and modern approaches toward resolving acoustical problems of large listening areas will be discussed.

7:55

2aAA2. Prediction of reverberation by stochastic processes in a forest: Creation of reverberation in an outdoor performance space. Toshiki Hanyu and Katsuaki Sekiguchi (Nihon Univ., Narashinodai, Funabashi, 274-8501, Chiba, Japan, hanyu@arch.jcn.nihon-u.ac.jp)

A forest is an outdoor space. For that reason, reverberation formulas of indoor spaces cannot be used when predicting forest reverberation. As a matter of fact, because the level decay curve of the reverberant sound produced in a forest is nonlinear, the concept of reverberation time is inapplicable. We clarified that reverberation of a forest is expressible as a Wiener process (Brownian motion), which is a stochastic process, and we derived the formula of reverberation of a forest based on that fact. In this study, Monte Carlo simulation was also carried out. Results clarified the following. (1) The energy decay curve of the reverberation sound of a forest becomes a straight line on the logarithmic scale of time and energy; its line gradient is -2. (2) At a point far from a sound source, the reverberant sound energy increases first and begins to decrease after some period. (3) When the mean-free path is short, the sound level of reverberation becomes high. Furthermore, we devised some methods of generating reverberation in outdoor space using the forest reverberation mechanism. For example, a proposed reverberation wall can produce reverberations using only one wall incorporating sound scatterers.

8:15

2aAA3. Large outdoor performance venues: Acoustical lessons learned. William J. Cavanaugh and K. Anthony Hoover (327F Boston Post Rd., Sudbury, MA 01776)

The Tweeter Center for the Performing Arts (formerly Great Woods) in Mansfield, MA, off Interstate 495, midway between Boston and Providence, accommodates 19 900 patrons and offers 75 to 85 performances each year between 15 May and 15 October. The Pittsburgh Symphony Orchestra inaugurated Tweeter Center's first concert season in June 1986, and The Center hosted the PSO as its summer home for three additional seasons. The Center has hosted performances from solo comedians to symphonic orchestras and rock bands, featuring such diverse acts as the Boston Pops, James Taylor, Phish, and The Kinks. The acoustical lessons learned over two decades of operation are described. The Center has served as a model for the design and operation of facilities in the 20 000-capacity range, including the often difficult issues of controlling excessive amplified sound transmission to residential neighbors, as well as assuring good listening conditions for patrons and performers alike.

2aAA4. Design of public address system for emergency evacuation in a tunnel. Sakae Yokoyama, Hideki Tachibana (Chiba Inst. of Technol., 2-17-1 Tsudanuma, Narashino-city, Chiba, 275-0016, Japan, sakae@iis.u-tokyo.ac.jp), Shinichi Sakamoto (Univ. of Tokyo, Meguro-ku, Tokyo, 153-8505, Japan), and Seiya Tazawa (Metropolitan Expressway Co. Ltd., Chiyoda-ku, Tokyo, 100-8930, Japan)

As a means to improve speech intelligibility of a public address PA system for emergency evacuation announcement in a tunnel, the authors have been investigating the application of the successive time-delay technique and have performed the experiments in an actual tunnel. In the experiments, the effect of the technique has been examined by comparing the difference of speech intelligibility between the conditions with/without time delay. For the design of the PA system in a reverberant tunnel, four kinds of psycho-acoustical experiments on speech intelligibility were performed in an anechoic room using the six-channel recording/reproduction technique and the following results have been obtained: (1) the successive time-delay technique is very effective to improve the speech intelligibility in a tunnel, (2) the system is considerably robust against the error in the delay time, (3) the speech intelligibility improves with the increase of the speech-rate but it should be suitable for such a tense situation as fire emergency, and (4) in order to further improve the speech intelligibility, it is necessary to control the reverberation in the tunnel by any sound absorption treatment. In this paper, the outline of the field experiment and the results of subjective experiments are introduced.

8:55

2aAA5. The Jay Pritzker Music Pavilion at Millennium Park, part 1: Design and construction. Richard Talaske (TALASKE, 105 N. Oak Park Ave., Oak Park, IL 60301)

The acoustic and audio system design for the Pritzker Pavilion represents a leap forward in the outdoor listening experience. The venue offers superior sound experiences to audiences of 11 000 using audio reinforcement and acoustic enhancement systems to support all types of musical performances, particularly for the resident Grant Park Music Festival orchestra. With reliance on the audio systems to project and envelop the audience in sound, the design of the stage enclosure was able to focus on the acoustic quality for the ensemble, a rare opportunity in outdoor facilities. The stage enclosure design includes variable acoustic devices and has an emphasis on intraensemble communication. This presentation will discuss the design of the stage enclosure, the trellis-supported audio systems, and the ways in which the facility is used by resident and visiting musical performers.

9:15

2aAA6. The Jay Pritzker Music Pavilion at Millennium Park, part 2: Research and testing. Gregory Miller, Richard Talaske, Byron Harrison, Evelyn Way (TALASKE, 105 N. Oak Park Ave., Oak Park, IL 60301), Lily Wang, David Bradley, and Craig Johnson (Univ. of Nebraska-Lincoln, Omaha, NE 68182-0681)

The design of the Pritzker Pavilion includes a number of innovative acoustic and audio elements. Two of these elements, the resilient orchestra riser system and the acoustic enhancement system, will be discussed in this presentation. Research previously presented to the ASA [Hoffman *et.al.*, 141st meeting of the ASA) was used in the design of the orchestra riser system at the Pritzker Pavilion. This presentation discusses the field tests that were conducted to assess vibration transfer across the surface of the riser system. Comparison is made to previous research and anticipated results. Subjective evaluations by orchestra members will also be discussed. The acoustic enhancement system is a specialized audio system, separate from the main reinforcement audio system, based on LARES digital signal processing and a distributed loudspeaker approach. Different settings have been developed over time to provide varying decay signatures that are employed on a music-program-dependent basis. Impulse response measurements were taken of unamplified sound, the reinforcement system, and the combination of the reinforcement system with the enhancement system. Testing was conducted with faculty and students from the University of Nebraska.

9:35

2aAA7. Acoustical design of the "Tenseishinbikai Hakko Shinden" Shrine in Shiga. Masayuki Ikegami (Dept. of Acoust Eng., Tech. Res. Inst., Obayashi Corp., 4-640 Simokiyoto, Kiyose City, Tokyo 208-8558, Japan)

A shrine called the "Tenseishinbikai Hakko Shinden" was designed and constructed. It covers a large area and has seating for 3690 people. It is used mainly for religious ceremonies and occasionally for concerts. This report outlines its acoustical design and noise control measures. The acoustics design specifically addressed reverberation control appropriate for its use, in addition to echo-nuisance prevention. The flat fan shape of the building often engenders insufficient initial sound reflection and generates echoes. To avoid these problems, in consultation with the designer, an irregular reflective surface was used for the side walls, a rib-shaped acoustic absorption surface, and a folding-screen shaped diffusion surface was used for the rear wall. In addition to conventional machine room noise control, particular attention had to be devoted to glass walls, rain noise, etc. For the glass walls, the effects on the outside (sound generated in the event room) and the effects on the inside (outdoor traffic noise) had to be controlled. To control rain noise, the sound generated by rain on the metal roof had to be predicted; a suitable roof material was selected.

9:55

2aAA8. Sound around the New England Patriots' stadia. Anthony K. Hoover, Andrew Carballeira, and Matthew Moore (Cavanaugh Tocci Assoc., Inc., 327 F Boston Post Rd., Sudbury, MA 01776)

The New England Patriots wanted to build a new football stadium in Foxboro, MA, adjacent to the old one which would be demolished after the new one was opened. The Environmental Impact Statement (EIS) indicated that one benefit would be a 3.34-dB reduction of stadium noise at the nearest town line because the distance would increase by 980 ft. A procedure to test the resultant noise reduction was developed with keen interest from the Massachusetts Department of Environmental Protection. Touring rock

concert sound systems would output high levels of pink noise, twice during sound-checks in the old stadium and then twice in the new stadium; resultant levels would be measured at several locations around the old stadium and then again around the new stadium. This paper will summarize the results of the measurements and will discuss interesting aspects of a test procedure designed by committee, including the importance of declining to be responsible for certain expectations of an EIS.

10:15–10:35   Break

10:35

2aAA9. Sound system design for acoustically challenging spaces.   Akira Mochimaru and Soichiro Hayashi   (Bose K.K., 28-3-Maruyama-Cho, Shibuya YT Bldg., Shibuya-Ku, Tokyo, Japan)

Sound reinforcement (SR) systems take a critical part in controlling acoustic performance in large venues. Loudspeaker system design including consideration of room acoustics becomes crucial in providing audio information with good intelligibility and natural tonal balance throughout the audience area. There are many elements that can spoil acoustic performance of the sound system in large spaces. Relatively long reverberation time, number of speaker clusters with distances in between, size of each speaker cluster that creates unexpected phase cancellations and so on. Neither the single speaker product nor sound system design based on experiences can provide a perfect solution for these kinds of spaces. On the other hand, an advanced acoustic simulation system including consideration of relative phase between sound sources with capability of listening to the simulated results is now available. By using this technology with academically proven accuracy, speaker array design methodology called bass-array was produced. Bass-array allows the sound system designer to control radiation directivity characteristics of low frequency. Design of custom line-array to suit room configuration of the space is another deliverable of the technology. This technology in association with bass-array or line-array has been successfully applied for providing appropriate acoustic solutions in a number of acoustically difficult spaces.

10:55

2aAA10. Recording in an outdoor performance space.   Pamela Harght   (Berklee College of Music, 150 Massachusetts Ave., Boston, MA 02115)

For both audio engineers and musicians, dealing with indoor and outdoor spaces poses several types of problems for recording and performance purposes. While recording indoors poses many problems, most of them (pertaining to room design, ventilation, placement of musicians) are controllable. However, dealing with outdoor spaces creates many issues entailing uncontrollable situations, specifically those entailing the environment. During the summer of 2006, the author spent 8 weeks at the Tanglewood Music Center in Lenox, MA on an audio engineering fellowship. Tanglewood is the summer home for the Boston Symphony Orchestra; it has both an outdoor performance space and a concert hall that opens for additional lawn seating. This paper discusses and compares the many issues of recording classical music outdoors as well as indoors, based upon experience, research, and interviews with musicians and members of the audio department.

11:15

2aAA11. Acoustics for the one percenter.   Scott Pfeiffer   (Threshold Acoust. LLC, 954 West Washington, 5th Fl., Chicago, IL 60607)

GMC adopted the one percenter from outlaw bikers who took pride in identifying themselves as the 1% of bikers that were making a bad name for the rest of the 99% that were law abiding. GMC and the outlaw bikers had the right idea. We all like to identify ourselves with the few, the proud, those that appreciate the special, the esoteric, the pursuit of greatness. Large venue projects present a special challenge to reaching these goals as the issues of getting 5000 or more people in and out begins to drive the design, natural acoustics become challenging or impossible, and the distribution of audio and video begins to depend on delay clusters and secondary distribution, which all begin as compromises. Doubt enters as the question is raised of how many patrons will notice. It is in these moments when finding the design that serves the one percenter makes all of the difference. Special examples of one percenters will be presented as a retrospective of large venues honoring the most successful projects in this type.

11:35

2aAA12. Improving sound qualities at the seats under deep balcony overhangs.   Wei-hwa Chiang, Wei Lin, and Kun-tai Wu (Dept. of Architecture, Natl. Taiwan Univ. of Sci. and Technol., #43, Sec.4, Keelung Rd., Taipei, 106, Taiwan, edchiang1224@hotmail.com)

Source-receiver distances in a large hall could be reduced by increasing the balcony overhangs, which, however, may cause insufficient acoustical energy for the under-balcony seats. Besides, the reduction in the sense of reverberation cannot be easily mitigated. In this research various design strategies, including utilizing the technique of a flying balcony, were examined in a 1550-seat hall with deep balcony overhangs. Both BEM simulation and scale model measurements were performed. Optimizing the profile of the under-balcony space and utilizing ceiling and side reflectors near the platform were effective in enhancing early reflection. This is especially meaningful for high-frequency string sounds that radiate against the ceiling. Opening the rear part of the balcony that formed a partially flying balcony could compensate late reflections. Preliminary scale modeling results showed that approximately 15% increase in early decay time can be achieved with carefully designed ceiling geometry when the opening area was only 7% of the overhang area.

Session 2aAB

Animal Bioacoustics: Fish Bioacoustics: A Tribute to Arthur Myrberg I

Arthur N. Popper, Cochair
Univ. of Maryland, Dept. of Biology, College Park, MD 20742

Richard R. Fay, Cochair
Loyola Univ. Chicago, Parmly Hearing Inst., 6525 N. Sheridan Rd., Chicago, IL 60626

Chair's Introduction—8:00

Invited Papers

8:05

2aAB1. A tribute to Art Myrberg. John H. Stadler (Natl. Marine Fisheries Service, 801 Lilly Rd. NE, Lacey, WA 98503) and Samuel L. Gruber (Univ. of Miami, Miami, FL 33149)

Arthur A. Myrberg, Jr. passed away in April, 2005, and will be remembered by many as a pioneer in the fields of fish bioacoustics and behavior. His distinguished career spanned more than 40 years, and saw him become a world-recognized expert in these fields, publishing more than 90 peer-reviewed articles, some of which are considered seminal. During more than 35 years as a professor of Marine Science at the Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, he was a well-respected and admired scientist, teacher, lecturer, mentor, and advisor to 24 graduate students. This presentation will trace Art's life from his days as a youth in Lulenecks, Illinois, through his college career and his work with Nobel laureate Konrad Lorenz at the Max Planck Institute, Seiweisen, Germany, culminating in his long career at the University of Miami.

8:25

2aAB2. A history of fish bioacoustic studies. Arthur N. Popper (Dept. of Biol., Univ. of Maryland, College Park, MD 20742, apopper@umd.edu), Richard R. Fay (Loyola Univ. of Chicago, Chicago, IL 60626), and William N. Tavolga (Mote Marine Lab., Sarasota, FL 34236)

Interest in fish sound production and hearing dates back to the writings of Aristotle (384–322 BCE) and Pliny the Elder (23–79 CE), both of whom noted not only that fish make sounds, but also speculated that fish must be able to hear despite their not having apparent ears. The fact that fish could hear was first demonstrated with the work of G. H. Parker in 1903 and later by Karl von Frisch. Studies on sound production were sporadic in the early part of the 20th century and often based on anecdotes until technology derived from antisubmarine warfare became available after World War II. This equipment enabled investigators to listen to fish and correlate sounds with behavior. It was not until the early 1960s that intensive investigations started on fish hearing. This paper will review the history of the study of fish bioacoustics and discuss the contributions of a large number of investigators including those of Arthur Myrberg. Indeed, Art's research stands as a landmark in the field, as well as a fine example of true scholarship. Furthermore, his charismatic presentations of his work were instrumental in stimulating the investigations of other scientists.

8:45

2aAB3. Art Myrberg's contribution to the study of hearing in fishes. Richard R. Fay (Parmly Hearing Inst. and Dept. of Psych., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, rfay@luc.edu)

Art Myrberg was trained to study behavior from an ethological perspective. When it came to studying the hearing of the vocal and communicating damselfish species of the genus *Eupomacentrus*, he surprisingly decided to use conditioning methods in combination with the best technology then known to produce an excellently designed animal psychoacoustic (laboratory) study that remains today, 26 years later, an exemplary model for students in this field [A. A. Myrberg and J. Y. Spires, J. Comp. Physiol. **140**, 135–144 (1980)]. First, Myrberg well understood that sensitivity to acoustic particle motion was possible for these species, so he produced or simulated sound in the far, free field using a long tube with sound absorbers at the opposite end, and measured both sound pressure and acoustic particle motion. He documented that *Eupomacentrus* was sensitive to particle motion at 100 Hz, but was sound-pressure-sensitive at higher frequencies. He also understood that ambient noise was an important issue in hearing studies, and he did all that he could to measure and reduce it. The result was a "V"-shaped audiogram with greatest sensitivity at 500 Hz, and almost no interspecific variation. We can have great confidence in these audiograms today as valid measurements of hearing ability.

9:05

2aAB4. Art Myrberg's contributions to fish acoustic communication. Michael L. Fine (Dept. Biol., Virginia Commonwealth Univ., Richmond, VA 23284-2012)

Arthur Myrberg was an ichthyologist and ethologist who trained with the Nobel laureate Konrad Lorenz. Art's work exceeded the boundaries of communication (a signal produced by an animal for its own benefit) to include related phenomena such as acoustic attraction and repulsion of sharks, effects of noise, and the creation of the term interception to describe a communication signal picked

up (e.g., intercepted) by another receiver for its own benefit. His work included a wide range of species (sharks, surgeonfish, damselfish, cichlids, gobies, and even invertebrates and marine mammals). He worked at sea and utilized diving for observation and experimental manipulation and then set up lab colonies for further observation and experimentation. What is perhaps his signal contribution to acoustic communication followed this paradigm in an investigation in four related *Eupomacentrus* species that produce similar chirps during a courtship dip. By playbacks of natural and altered sounds he and his students separated the subtle cues used by these fishes to permit intraspecific communication. This rather dry abstract fails to capture Art's love of science, his ability to make outstanding presentations, and my value of his friendship.

9:25

2aAB5. Recent advances in fish acoustic behavior. Timothy Tricas (Dept. of Zoology and Hawaii Inst. of Marine Biol., Univ. of Hawaii at Manoa, Honolulu, HI 96822, tricas@hawaii.edu)

Sounds produced by fish provide important signals or cues in many behavioral contexts. Mechanisms of sound production are diverse and best described in fishes with specialized anatomical structures, but remain tentative for many species. The coproduction of low-frequency hydrodynamic and acoustic stimuli during behavioral interactions can excite the inner ear, lateral line, or both to provide complex acoustico-lateralis information. Recent behavior studies show that acoustic signals function in behavioral displays of aggression, species recognition, courtship, mate attraction, and spawning; in some cases it can function as a single modality signal. Sound stimuli often overlap with the hearing band of highest sensitivity (100–1000 Hz for most species), but in some species these do not match. Much data now exist on the diversity of sounds produced by fishes and environmental acoustic cues, but more evaluation is needed within the context of the animals natural environment. Examples include acoustic cues for recruitment of pelagic fish larvae to reefs, use of ultrasound to avoid predators, and infrasound stimuli. Comparative work on both sound production and sensory reception in fishes can be assessed in relation to their ecology to better understand the evolution of fish acoustic behavior.

9:45

2aAB6. Lateral line research: Recent advances and new opportunities. Sheryl Coombs (Dept. of Biological Sci., Bowling Green State Univ., Bowling Green, OH 43403) and Horst Bleckmann (Univ. of Bonn, 53115 Bonn, Germany)

Computational fluid dynamics, digital particle imaging velocimetry, and engineered micro flow sensors, along with emerging collaborations between neurobiologists and engineers, have opened up new vistas for lateral line research in recent years. Engineering advances have increased our ability to measure the spatiotemporal characteristics of biologically significant, more complex flow signals and how these signals are sensed and processed by the lateral line. Following in the rich tradition of Art Myrberg's work on sound communication in fishes, studies on hydrodynamic signal production and use by fish during interactive social behaviors are now feasible. In addition to these new opportunities, several recent studies have enriched the lateral line landscape. The lateral line sensing and encoding of vortex trails shed by swimming fish or obstacles in a stream have been investigated, along with associated behaviors (e.g., vortex trail following and obstacle entrainment). Blind Mexican cavefish have been shown to form cognitive maps of their surroundings using lateral-line based, hydrodynamic cues. Finally, the responses of goldfish auditory and lateral line systems to the same stimulus source have been documented to reveal how these two closely related systems might work in concert to extract complementary but different types of information from nearby hydroacoustic sources.

10:05–10:20   Break

Contributed Papers

10:20

2aAB7. Year-round changes of ambient noise in freshwater habitats and temperature effects on hearing in fishes. Sonja Amoser and Friedrich Ladich (Dept. of Neurobiology and Behavioral Sci., Univ. of Vienna, Althanstr. 14, 1090 Vienna, Austria)

Fish get crucial information for their survival from their acoustic environment. Therefore, changes in ambient noise as well as temperature over the year potentially impose great challenges for fishes, especially in freshwater habitats in temperate climates. Noise levels vary more in stagnant habitats such as lakes, backwaters, and streams than in large rivers. The spectral profiles change in stagnant habitats due to biotic noise sources in summer (e.g., stridulating insects), whereas in small streams spectral composition differs due to small changes in water volume, e.g., after snow melt. To investigate possible seasonal effects on fishes, we studied the hearing capabilities of a hearing specialist (Common carp) and a hearing nonspecialist (European perch) at two different temperatures (12 °C and 22 °C) using the AEP recording technique. In the carp the hearing sensitivity decreased up to 12 dB (mean decrease: 4 dB) at 12 °C, whereas no such change was observed in the perch. Latencies of auditory evoked potentials became longer by approximately 0.8 ms at lower temperatures in the carp. The results indicate that auditory sensitivity in hearing specialists is affected by changes in temperature as well as ambient noise throughout the year. [Work supported by FWF.]

10:35

2aAB8. The influence of ambient temperature and thermal acclimation on hearing in an eurythermal hearing specialized fish. Lidia E. Wysocki and Arthur N. Popper (Dept. of Biol., Univ. of Maryland, College Park, MD 20742, lwysocki@umd.edu)

Many fish species live in habitats whose water temperature underlies seasonal and diurnal fluctuations. Being ectothermic animals, their body temperature depends on the temperature of the ambient water. Since the speed of all metabolic processes depends on body temperature, the ambient temperature might affect various physiological processes including all sensory systems. However, the maintenance of sensory functions in a changing environment can be crucial for an animal's survival. Several fish species strongly rely on their auditory abilities for acoustic orientation, communication, detecting predators and prey, and finding mates. Channel catfish (*Ictalurus punctatus*) was chosen as a model for an eurytherm species with a relatively high hearing sensitivity and bandwidth to investigate the influence of ambient temperature and thermal acclimation on the auditory system. Hearing sensitivity was measured at different water temperatures to which the animals were either unacclimated or acclimated by recording auditory-evoked potentials. The ambient water temperature significantly influenced hearing thresholds and the shape of the auditory-evoked potentials, especially at higher frequencies. Hearing sensitivity remained decreased at low water temperatures even after an acclimation period of one month, indicating that certain physical limits on sensory processes cannot be compensated for by thermal acclimation.

10:50

2aAB9. Investigating hearing in sturgeon: Getting closer to the origin of hearing in vertebrates. Michaela Meyer, Arthur N. Popper (Dept. of Biol., Univ. of Maryland, College Park, MD 20742), and Richard R. Fay (Loyola Univ. of Chicago, Chicago, IL 60626)

It has been hypothesized that hearing arose early in vertebrate evolution to use sound to glean information about the environment, including the presence of predators and prey. The importance of sound to sense the acoustic scene is especially true underwater where the visual horizon is limited. To create an acoustic scene from sound, an animal needs to localize and identify sound sources. We hypothesize that the sound localization capabilities of primitive bony fishes such as sturgeon are as capable as those of modern bony fishes such as goldfish and, perhaps, many terrestrial vertebrates. We present results from experiments investigating the neural mechanisms of directional hearing and frequency analysis in the auditory periphery of the sturgeon. Extracellular single units were recorded from the eighth nerve during stimulation with a three-axis shaker system that subjects the fish ear to linear accelerations in known directions at various frequencies. This system enables the characterization of both spatial and frequency tuning of the afferents innervating the ear. The results show that similar neural mechanisms for encoding sound direction and frequency exist in the periphery of modern and primitive bony fish, thereby supporting the hypothesis that basic auditory functions arose very early in vertebrate history.

11:05

2aAB10. Recent advances in the fish auditory efferent system. Zhongmin Lu (Dept. of Biol., Univ. of Miami, 1301 Memorial Dr., Coral Gables, FL 33146, zlu@miami.edu) and Seth Tomchik (Univ. of Miami, Miami, FL 33101)

The octavolateral efferent neurons in fish project axons to innervate the ear and lateral line organs. However, the function of the octavolateral efferent system is not well understood. This review focuses on recent progress in structure and function of octavolateral efferent system in a teleost, the sleeper goby (*Dormitator latifrons*). The sleeper goby has a single octavolateral efferent nucleus that is composed of about 19 large, densely packed somata in the medulla. Most efferent neurons have bilateral dendrites and single axons innervating ipsilateral peripheral octavolateral organs. The efferent neurons have distinctive characteristics of frequency, temporal, directional, and dynamic responses, which differ from those of otolithic organ afferents. How efferent stimulation affects afferent response is investigated by electrically stimulating efferent somata and recording compound action potentials from afferents innervating the saccule, the main auditory organ in this species. There are different efferent modulatory effects on afferent response, depending on the level of background noise. Efferent activation inhibits afferent response in quiet conditions but can enhance signal detection in noisy environments. Results of these studies indicate that the fundamental role in efferent modulation of afferent input is likely conserved in both nonmammalian vertebrates and mammals. [Work supported by NIH/NIDCD Grant R01DC003275.]

11:20

2aAB11. Flows near model otoliths and their implications for fish hearing. Charlotte W. Kotas, Peter H. Rogers, and Minami Yoda (Mech. Eng. Dept., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The fish inner ear contains a dense, bony otolith that oscillates in response to acoustic excitation. This relative motion between the otolith and its surrounding fluid generates both steady streaming and unsteady flows in the fluid which in turn displaces the hair cells on the macular membrane. Fish "hear" by interpreting these hair cell displacements. However, most of the hair cells are located near a groove along the otolith (the sulcus), making it likely that the irregular geometric features of the otolith will affect how fish hear by influencing the motions of these hair cells. To understand these acoustically induced flows, magnified otolith models were constructed based upon three-dimensional computer tomography scans of actual otoliths. The model otoliths were sinusoidally oscillated so that steady streaming and unsteady flows could be experimentally studied using particle-image velocimetry. The effect of geometric features like the sulcus was studied near the model surface at various oscillation frequencies, ω, and amplitudes, s. Results will be presented for Reynolds number $Re \equiv \omega L^2/\nu \approx 1 - 10^2$ and amplitude ratios $\epsilon \equiv s/L = 0.05 - -0.2$, where L is a representative otolith length scale and ν is the kinematic viscosity of the fluid. [Work supported by ONR.]

11:35

2aAB12. Particle motion perception in cephalopods. Kenzo Kaifu, Susumu Segawa (Tokyo Univ. of Marine Sci. and Technol. 4-5-7, Kounan, Minato-ku, Tokyo, Japan), and Tomonari Akamatsu (Fisheries Res. Agency, Kamisu, Ibaraki, Japan)

In the 1980s, there still had been the argument whether cephalopods could detect sound stimulus or not. The ability of sound detecting (hearing ability or an ability to detect vibration) in cephalopods was proved around 1990, however, there have been few studies about sound detection in cephalopods afterwards. Based on morphological and behavioral studies, it seems that cephalopods do not perceive sound pressure. Provided that compressional waves do not stimulate the nervous systems of the animals, they do response to sound waves due to water particle movement. In this study, the ability to detect sound waves in cephalopods was investigated from the standpoint of water particle movement perception. An object would be well vibrated by sound waves when the length of the object is much smaller relative to the wavelength; consequently, when an animal detects water particle movement, the smaller the animal is, the higher frequency and weaker stimulus it would detect.

Session 2aAOa

Acoustical Oceanography and Underwater Acoustics: Acoustic Scattering by Aquatic Organisms III

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Masahiko Furusawa, Cochair
Tokyo Univ. of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo 108-8477, Japan

Chair's Introduction—8:00

Invited Papers

8:05

2aAOa1. Fish population and behavior revealed by instantaneous continental-shelf scale imaging. Nicholas C. Makris (Dept. of Mech. Eng., MIT, 77 Massachusetts Ave., Cambridge MA 02139), Purnima Ratilal (Dept. of Elec. and Comput. Eng., Northeastern Univ.), Deanelle T. Symonds, Srinivasan Jagannathan, Sunwoong Lee (Dept. of Mech. Eng., MIT, Cambridge, MA), and Redwood W. Nero (Naval Res. Lab., Stennis Space Ctr., MS)

Fish populations in continental shelf environments are instantaneously imaged over thousands of square kilometers and continuously monitored by a remote sensing technique in which the ocean acts as an acoustic waveguide [Science **311**, 660–663 (2006)]. The technique, ocean acoustic waveguide remote sensing (OAWRS), has revealed the instantaneous horizontal structural characteristics and volatile short-term behavior of very large fish shoals, containing tens of millions of fish and stretching for many kilometers. The former follows a fractal or power-law spectral process, indicative of structural similarity at all scales. The latter features compressional waves of fish population density that travel roughly an order of magnitude faster than fish can swim. Here we present an overview of OAWRS imagery documenting fish activity over a 2-week period in the Spring of 2003 at the edge of the continental shelf, roughly 200 km south of Long Island, NY. Until now, continental shelf environments have been monitored with highly localized line-transect methods from slow-moving research vessels. These methods significantly undersample fish populations in time and space, leaving an incomplete and ambiguous record of abundance and behavior.

8:25

2aAOa2. Acoustical imaging of nonbladdered animals using underwater acoustic camera. Kohji Iida, Rika Takahashi, Tohru Mukai (Fac. Fish., Hokkaido Univ., 3-1-1 Minato, Hakodate, Hokkaido, 0418611 Japan, iidacs@fish.hokudai.ac.jp), Yong Tang (Dalian Fish. Univ., Dalian 116023 China), and Masanori Sato (Honda Electron. Co., Ltd., Toyohashi, Aichi, 4413193 Japan)

An underwater acoustic camera using 3.5-MHz ultrasonic waves enclosed in a pressure-resistant case was constructed to observe underwater marine animals. This enabled the measurement of the size, shape, and behavior of living marine animals including internal organs by B-mode images in the detection range up to 240 cm. Observations were conducted for captive animals in a water tank and for natural animals in a field. The captive animals, including fish, squid, and jellyfish, were observed, and a three-dimensional internal structure of animals was reconstructed using multiple acoustical images. The most important contributors of acoustic scattering were the vertebra for nonbladdered fish, and the liver and reproductive organs for invertebrate animals. In a field experiment, the shape, size, and swimming behavior of wild animals were observed. The possibilities and limitations of the underwater acoustic camera for fishery applications were discussed. [Work supported by JSPS Grant-In-Aid for Scientific Research.]

8:45

2aAOa3. Simultaneous observations of zooplankton distribution and density using net sampling, a laser-optical plankton counter, and acoustics. Tohru Mukai, Yasuzumi Fujimori, Jun-ichiro Kishi, and Kohji Iida (Faculty of Fisheries Sci., Hokkaido Univ., 3-1-1 Minato, Hakodate, Hokkaido 0418611, Japan)

Conventional plankton sampling methods are based on net tows, which are often poorly suited to capturing spatiotemporal variability. Various automatic plankton samplers based on optical principles have been developed recently. The laser-optical plankton counter (LOPC) was developed as the next generation OPC. In addition, acoustic technology has been used extensively to survey zooplankton distribution and abundance. This study compared the density and size spectra determined using net contents, LOPC data, and acoustic data. Surveys were conducted for 3 years around Hokkaido, Japan. The sound-scattering layer (SSL) was recorded on the echogram of an echosounder. Biological sampling was conducted using a 1-m^2 frame-type mid-water trawl (FMT; mesh size 1.5 mm) and a NORPAC net (mesh size 500 microns). An LOPC was mounted inside the mouth of these nets. Net sampling was carried out at dusk, when the SSL moves toward the surface (ca. 10–35 m). Acoustic data were derived using a Simrad EK60 echosounder operating at 38, 120, and 200 kHz simultaneously. A comparison of the use of acoustic data, LOPC data, and net contents to assess zooplankton distribution and density will be presented and discussed. [Work supported by JSPS Grants-In-Aid for Scientific Research.]

9:05

2aAOa4. Riding the wave: Scale-dependent descriptions of walleye pollock distributions. Julian M. Burgos, John K. Horne (School of Aquatic and Fishery Sci., Univ. of Washington, 1122 NE Boat St., Seattle, WA 98105, jmburgos@u.washington.edu), and John K. Horne (Natl. Marine Fisheries Service, NOAA, Seattle, WA 98115)

Wavelet analysis was used to describe scale-dependent distributional variability of walleye pollock (*Theragra chalcogramma*) in the eastern Bering Sea shelf between 2000 and 2002. Area backscattering strength (ABC), a measure of acoustic density, was calculated at 20-m horizontal intervals in five echo-integration trawl surveys. The maximum overlap discrete wavelet transform was used to partition the ABC variance into scales ranging from 0.04 to 10.24 km. Wavelet variance was estimated for each transect after denoising with phase plane thresholding. Wavelet spectra can be characterized by large variances at small scales (less than 0.64 km), decreasing when scale increased to 1.28 km, and increasing variance at larger scales. The fraction of the total variance assigned to small scales was negatively correlated with the total acoustic density. This is attributed to the aggregative behavior of walleye pollock. Peaks in wavelet spectra, indicating the presence of coherent features (i.e., schools or shoals), were observed in 48% of the transects at scales ranging from 1.28 to 10.24 km. Peaks at larger scales were more common in winter transects, when walleye pollock form large spawning aggregations.

9:20

2aAOa5. An overview of the theoretical framework for measuring fish swimming speed by a sector-scanning sonar. Akira Hamano (Natl. Fisheries Univ., 2-7-1, Nagata-honmachi, Shimonoseki, Japan 759-6595, ahamano@fish-u.ac.jp), Jennifer Miksis-Olds, Christopher Jakubiak, and Ernest Bernard (Univ. of Massachusetts Dartmouth, New Bedford, MA 02744-1221)

Two identical high-frequency sector-scanning sonars, mounted on a stationary acoustical-optical platform (AOP), were used to study the swimming speed of fishes. The first phase of this study consisted of a motion calibration performed in an acoustic test tank using a standard target sphere (WC 38.1 mm). The sonars are a 200-kHz sector-scanning system employing an 8-element array to receive echo signals over a 90° arc. An overview of the theoretical framework for the motion calibration of the sector-scanning sonar is presented. The second phase of the study related the calibration data to the observed movement of fish. The two-dimensional swimming tracks were identified from the digitized sonar data recorded while the AOP was deployed on the seafloor. These tracks were compared with the movement of the target sphere to make estimates of fish swimming speed. Optical data, when available, were used to confirm the results of acoustic data. The limitations and precision of swimming speed measurements using this system are discussed. The application of this methodology to studies of fine-scale fish behavior has the potential to make previously intractable questions possible. [Work supported by NOAA.]

9:35

2aAOa6. Depth- and size-dependent rhythms of animal emergence in an estuary. Mei Sato and Peter A. Jumars (Darling Marine Ctr., Univ. of Maine, 193 Clark's Cove Rd., Walpole, ME 04573, mei_sato@umit.maine.edu)

Among migrating crustaceans, some species inhabiting shallow waters stay very close to the bottom or burrow into the sediments during the day and come up to the surface at night. This behavior, called emergence, was monitored by the Tracor Acoustic Profiling System (TAPS) with six frequencies (265–3000 kHz) to clarify the mechanisms of benthic-pelagic coupling. Instead of one emergence event up to the surface at night, data with high spatial resolution indicate multiple emergence events varying with both the size of the organisms and distance above the bottom, suggesting that for some species there is a target depth beyond which upward migration is limited. In the Damariscotta River estuary, ME, nocturnal emergence events of mysid shrimp, *Neomysis americana*, occur from late June through early November. Mysid shrimp are known to be important food sources for fishes such as longhorn sculpins, smelt, alewives, and Atlantic cod. The timing of their emergence near the bottom corresponds to the maximum flood tide, suggesting enhancing retention in the estuary. Smaller organisms, on the other hand, emerge primarily during periods of slack tide, but reach the surface only at slack low, also avoiding washout.

9:50

2aAOa7. Combining acoustics and video to image fish in the acoustic dead zone. Jennifer L. Miksis-Olds, Ernest Bernard, Christopher J. Jakubiak, and D. V. Holliday (School for Marine Sci. and Technol., Univ. of Massachusetts, 838 S. Rodney French Blvd., New Bedford, MA 02744, jmiksis@umassd.edu)

A stationary acoustical-optical platform (AOP) has been developed to detect groundfish presence and to estimate biomass. The AOP is a new approach to fisheries science because sampling from a platform on the ocean floor overcomes limitations of traditional surface acoustic surveys, where echoes from groundfish near the seafloor are obscured by stronger reflections from the ocean bottom. The AOP capitalizes on the complementary strengths of acoustical and optical sampling techniques by using acoustics for fish enumeration and optics for species identification and measurements of fish length. Each AOP is equipped with four video cameras, two 200-kHz sonar transducers, and a suite of environmental sensors. One sonar face is mounted in the horizontal direction to image the area directly above the seafloor, whereas the second transducer is oriented vertically to sample the water column above the platform. Acoustic backscatter, target species, size, and observation angle were quantified when fish were simultaneously detected by the sonar and optical sensors. Calculations of target strengths were then related to fish size and orientation, for incorporation into biomass estimates. The AOP is designed to be used in an observation system with multiple platforms elements, so fish populations can be monitored over large spatial scales. [Work supported by NOAA.]

10:05–10:20   Break

10:20

2aAOa8. Acoustic and optic characterization of forage fish distribution in the Bering Sea. Sandra Parker Stetter, John Horne (School of Aquatic and Fishery Sci., Univ. of Washington, Seattle, WA 98195-5020), and James Churnside (Natl. Oceanograph. and Atmospheric Administration Environ. Technol. Lab., Boulder, CO 80305)

Mesopelagic forage fish species are important components of the Bering Sea ecosystem, but information on species distribution and identity is limited. Using acoustics and optics [LIght Detection And Ranging (LIDAR)], this study compared distribution patterns detected by the two gears. Forage species distribution and quantity differed between shelf (6–100 m) and slope (6–100 m, 100–300 m, 300 m to bottom) regions. Acoustic results suggested that shallow and deep depth zones contained dispersed backscatter while the middle slope layer contained patchy schools associated with the shelf break. Variogram results for repeated LIDAR surveys of the shelf and slope regions indicated that backscatter distribution between 6 and 30 m was dynamic at the scale of days. This result was expected given the strong frontal nature of the area. When LIDAR results were compared with coincident acoustic transects on the shelf and slope, differences were found in gear detection of backscatter. Based on acoustic results, 25%–63% of forage fish in the shelf and slope regions were deeper than the LIDAR detection range. Although both LIDAR and acoustics are constrained to portions of the water column, the utility of remote sampling technologies is dependent on survey objectives. [Funding was provided by the North Pacific Research Board.]

2a WED. AM

10:35

2aAOa9. Comparing repeated sidescan sonar measurements of benthic squid egg beds in Monterey Bay, CA. Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Pat J. Iampietro, and Mary Young (California State Univ., Monterey Bay, Seaside, CA 93955)

The squid *Loligo opalescens* lays its eggs in capsules attached to a flat sandy substrate. These capsules are usually clustered in so-called egg beds. Earlier work in Monterey Bay, CA [Foote *et al.*, J. Acoust. Soc. Am. **119**, 844–856 (2006)] established the acoustic detectability of such egg beds. This work has now been extended in a new study performed in June 2006. An EdgeTech sidescan sonar was towed at nominal 5-m height over the bottom. Measurements were made at 400 kHz, horizontal beamwidth 0.5 deg, along six parallel transects, with adjacent centerlines separated by 30 m. The swath width along each transect was 100 m. The transects were repeated on 10, 13, and 14 June. Data from areas where egg beds were densely concentrated, confirmed by drop video camera, have been excerpted. Comparisons of corresponding data, which are generally very strong, are quantified. The question is asked as to whether such sidescan sonar data can be used differentially to measure egg-laying success or hatching over rather short periods of time. If so, this could be a useful tool both in ecological studies and for management purposes. [Work supported by NURP.]

10:50

2aAOa10. Estimating pressure gradients governing fish shoal dynamics over continental-shelf scales. Srinivasan Jagannathan, Sunwoong Lee, Deanelle T. Symonds (Dept. of Mech. Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 02139), Purnima Ratilal (Northeastern Univ.), and Nicholas C. Makris (MIT, Cambridge, MA 02139)

The environmental pressure gradient and stress field on fish shoals spanning more than 10 km is calculated from a sequence of instantaneous continental-shelf scale OAWRS images of areal fish population density [Science **311**, 660–663 (2006)]. The calculation is based on continuum

mechanics and the velocity field of fish groups previously estimated from OAWRS imagery using acoustic flow [Jagannathan *et al.*, J. Acoust. Soc. Am. **119**, 3435 (2006)]. These environmental stress fields describe the driving mechanism behind the dynamic behavior of large fish groups such as expansion, contraction, and translation. Regions of high stress are correlated with regions of high areal fish density. They are also correlated with density waves that convey information across a fish group. Our estimates of environmental stress on fish groups can be correlated with environmetal parameters such as plankton, temperature, and current distributions through physical and biological driving mechanisms.

11:05

2aAOa11. Sonar mapping for coral reef conservation. Lisa Zurk, Joe Lotz, Tim Ellis, James McNames (Elec. and Comput. Eng. Dept., Portland State Univ., P.O. Box 751-ECE, Portland, OR 97207, zurkl@cecs.pdx.edu), and Jean-Louis Ecochard (The Nature Conservancy, Arlington, VA 22203)

Many conservation efforts require an inexpensive and easy-to-use method of mapping and monitoring coral reef structure. A desirable product is a 3D map that effectively portrays the reef relief (and potentially the sub-structure) or features such as fish locations and fish spawning areas. Given field site limitations associated with the reefs, the mapping technique should be portable, rugged, and easy to use. One solution, developed at the Nature Conservancy (TNC), is to utilize a Lowrance fish finding sonar with a GPS system to estimate the top of the coral. This approach has been employed for TNC mapping exercises to generate ArcGIS maps of conservation areas. However, these maps are often negatively affected by sampling and processing artifacts such as sparse collection geometries (causing interpolation errors), distortion due to beam shape, and changes in boat (transducer) position due to wave motion. In this presentation we discuss recent work to address these effects, as well as enhancements to the system by processing acoustic returns from the coral substructure or within the water column for detection of fish populations. The presentation will provide a description of future research, including the intent of making the finished system available as a conservation tool.

Invited Papers

11:20

2aAOa12. Robust analysis of single echo trace obtained by split beam sounder. Masahiko Furusawa (Tokyo Univ. of Marine Sci. and Technol., 4-5-7 Konan, Minato Tokyo 108-8477 Japan, frsw@s.kaiyodai.ac.jp) and Kazuo Amakasu (Natl. Res. Inst. of Fisheries Eng., Kamisu Ibaraki 314-0408 Japan)

The theory and method of the echo trace analysis of single echoes (ETA), which simultaneously analyze the shape and level of single echoes, are almost established. But, the method has not been extensively used. The major reasons are high-quality echoes with a high signal-to-noise ratio (SNR) necessary for the analysis are some times difficult to obtain; the error introduced by transducer motion is large; and conditions for the theory sometimes could not be fulfilled. Therefore, we apply the ETA to the split beam method, which is most advantageous for the ETA, and make the method robust by introducing the following: careful echo selection considering SNR, correction of the effect of transducer motion, smoothing of the trace by regression analyses, introduction of indexes checking reliability, etc. The useful results of our ETA are the track, the swimming speed and orientation, the target strength as a function of tilt angle, and the estimated size of individual fish. The robust ETA was applied to echoes of fish near a sea bottom and of Antarctic krill to confirm the effectiveness of the method.

11:40

2aAOa13. Comparing acoustic model estimations and *in situ* target strength measurements of juvenile walleye pollock (*Theragra chalcogramma*). Kazuhiro Sadayasu (Natl. Res. Inst. of Fisheries Eng., Fisheries Res. Agency, 7620-7 Hasaki, Kamisu, Ibaraki 314-0408, Japan), Tohru Mukai, and Kohji Iida (Hokkaido Univ., Hakodate, Hokkaido 041-8611, Japan)

Recently, many acoustic surveys for the juvenile walleye pollock have been done around the Hokkaido, Japan. Dependence of target strength on fish length is a most important factor in the conversion of acoustic data to fish biomass. However, the relation between target strength and length of juvenile walleye pollock is not well investigated. This study examined the target strength length relationship (3–10 cm) for the echo integration measurements of the juvenile walleye pollock at 38 kHz from acoustic theoretical model predictions and *in situ* measurements. A prolate spheroid model was used to compute the target strength of the fish relative to its orientation. Swimbladder morphologies of live juvenile walleye pollock caught by set net were obtained from transmission of light or soft x-ray, and converted their morphological parameters. Biological samplings using framed mid-water trawl were carried out near

the spawning area of walleye pollock. During fish samplings, *in situ* target strength measurements were conducted using quantitative echo sounder with split beam transducer. While predicted mean target strengths matched mean *in situ* measurements at small length, there was not agreement at the large length. This discrepancy is a result of avoidance of large fishes from a net mouth.

WEDNESDAY MORNING, 29 NOVEMBER 2006

KAUAI ROOM, 7:45 A.M. TO 12:05 P.M.

Session 2aAOb

Acoustical Oceanography and Signal Processing in Acoustics: Acoustic Tomography for Coastal and Deep Water Applications III

James A. Mercer, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Arata Kaneko, Cochair
Hiroshima Univ., Graduate School of Engineering, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan

Chair's Introduction—7:45

Contributed Papers

7:50

2aAOb1. Moving ship thermometry. James A. Mercer, Nick Alger, Bruce M. Howe, Rex K. Andrew (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, mercer@apl.washington.edu), and John A. Colosi (Naval Postgraduate School, Monterey, CA 93943)

The long-range ocean acoustic propagation experiment (LOAPEX), conducted in the NE Pacific Ocean, provided acoustic transmissions from a ship-suspended source at seven stations spanning a near zonal path of 3200 km. The transmissions were received on several bottom-mounted horizontal hydrophone arrays distributed about the NE Pacific Ocean Basin. The experiment and a simple inverse method are described and a preliminary thermal "snapshot" of the ocean basin is presented. [Work supported by ONR.]

8:05

2aAOb2. Anisotropy in the spatial distribution of the arrival points of a ray after a fixed travel time. Stanley M. Flatté (Phys. Dept., Univ. California Santa Cruz, Santa Cruz, CA 95064, sflatte@ucsc.edu) and John A. Colosi (Naval Postgraduate School, Monterey, CA 93943)

An acoustic pulse emitted from a point source arrives on a well-defined wavefront after a fixed travel time. In the geometrical-acoustics approximation, this behavior is understood by following rays. Two simple ray models are presented to treat the scattering of rays by internal waves, and to evaluate the ratio of movements in the ray end points within and perpendicular to the wavefront. Assuming weak forward scattering, and no waveguide, the model predicts that the end points move along the wavefront over much larger distances than perpendicular to the wavefront. [Work supported by ONR.]

8:20

2aAOb3. Fresnel zones in inhomogeneous media. Irina I. Rypina and Michael G. Brown (Univ. of Miami, RSMAS-AMP, 4600 Rickenbacker Cswy., Miami, FL 33149)

A simple general expression for the width of a Fresnel zone in a smooth inhomogeneous environment is derived. In a stratified environment at long range, the general expression is shown to reduce approximately to one that is proportional to $\sqrt{|\alpha| r(R-r)/f}$, where α is the ray stability parameter, f is the acoustic frequency, r is the range to the field point of interest, and R is the source-to-receiver range. In a stratified environment on which a weak perturbation is superimposed, at long range

the width of a bundle of micromultipaths is shown to be approximately proportional to $|\alpha|\sqrt{r(R-r)}$. Thus, at long range both the diffractive and scattering contributions to "the width of a ray" are proportional to $\sqrt{r(R-r)}$ and increase with increasing $|\alpha|$. Theoretical predictions are shown to agree well with travel time sensitivity kernel calculations. [Work supported by ONR.]

8:35

2aAOb4. Moving ship tomography from the Basin Acoustic Seamount Scattering Experiment (BASSEX-04). Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039), Arthur B. Baggeroer (MIT, Cambridge, MA 02139), Peter F. Worcester, and Matthew A. Dzieciuch (Univ. of California at San Diego, La Jolla, CA 92093)

During the Basin Acoustic Seamount Scattering EXperiment (BASSEX-2004) a set of 32 acoustic transmissions of four moored sources in the central Pacific were made. Broadband transmissions (100 Hz bandwidth, 250 Hz center frequency, 12.276 s period length) from two sources suspended (800 m depth) were received on a towed array (350 m depth, 200 m aperture) as it passed between the two sources at a range of 250 km. This data set provides the opportunity for examining the feasibility of underway moving ship tomography (MST). The primary oceanographic feature is the variability of the mixed layer over a 500×500-km^2 region. Accurate ocean acoustic tomography requires the development of an $N \times 2D$ inversion scheme as well as techniques to perform array element localization to within 10 m. Results on applying a Kalman-filter-based technique to the array sensor data (depth, pitch, roll) and the measured ADCP will be presented. With the receiving ship transecting the path between the two source moorings, each transmission has two paths that pass through different bodies of water.

8:50

2aAOb5. Vertical coherence of low-frequency sound waves propagating through a fluctuating ocean. Alexander G. Voronovich and Vladimir E. Ostashev (NOAA, Earth System Res. Lab., PSD4, 325 Broadway, Boulder, CO 80305)

Low-frequency sound waves propagating over large distances in the ocean exhibit significant scattering due to internal waves. Studies of low-frequency sound propagation through a fluctuating ocean are important for many practical concerns such as communication, acoustic tomography, and source detection, ranging, and recognition. In a recent work [A. G. Voronovich and V. E. Ostashev, J. Acoust. Soc. Am. **63**, 1406–1419

[(2006)] a 3-D modal theory of sound propagation through a fluctuating ocean was developed. Closed equations for the first two statistical moments of a sound field were derived using the Chernov method and assuming that the spectrum of internal waves is given by the Garret-Munk spectrum. In the present paper, we numerically solve the derived equation for the second moment. The results obtained are used to study the dependence of the vertical coherence of a sound field on several parameters of the problem: range, frequency, source and receiver depths, and parameters that characterize the Garret-Munk spectrum. [Work supported by ONR.]

9:05

2aAOb6. A partial theory of phase coherence loss in deep ocean propagation. Francisco J. Beron-Vera and Michael G. Brown (RSMAS-AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

In a fluctuating ocean, coherently adding multiple receptions of repeated transmissions of a signal generally results in a processing gain relative to the background noise only for some finite duration owing to a gradual loss of wave-field phase coherence. In this talk a theoretical framework for understanding some elements of phase coherence loss in transient wave fields in deep ocean environments is presented. The theory predicts that phase coherence loss is pathdependent and very different at short range than at long range. PE simulations of transient wave-field phase coherence loss in an idealized time-varying ocean are presented and shown to be in good qualitative agreement with theoretical predictions. [Work supported by ONR.]

9:20

2aAOb7. Vertical structure and arrival times of the axial wave based upon conductivity, temperature, and depth data acquired during the Long-range Ocean Acoustic Propagation Experiment. Natalie S. Grigorieva and Gregory M. Fridman (Dept. of Appl. Math., St. Petersburg State Marine Tech. Univ., 3 Lotsmanskaya Str., St. Petersburg, 190008, Russia, nsgrig@natalie.spb.su)

In the LOAPEX experiment at some distances from VLA the sound source was deployed to depths quite close to the sound-channel axis. This took place at stations T1000, T500, T250, and T50 where, in particular, the source was deployed to 800 m and the depth of the sound-channel axis at these ranges is about 680 m. The interference of the wave fields corresponding to near-axial rays leads to formation of a complex interference wave ("the axial wave") that propagates along the axis being a component of the axial crescendo. The talk will present the results on simulation of the axial wave as a function of depth of the receiver and the calculation of its arrival time at frequencies 56.25–93.75 Hz. In the time domain the pulse propagation modeling is carried out for the broadband acoustic signals with the pulse center frequency of 75- and 37.5-Hz bandwidth. The obtained results are compared with those for the lowest mode in the adiabatic approximation. Calculations are made for the range-dependent ocean model based upon LOAPEX CTD data and use the integral representation of the axial wave in a local coordinate system introduced in the vicinity of the range-variable sound-channel axis. [Work supported by ONR.]

9:35

2aAOb8. Evolution of a downslope propagating acoustic pulse. Kevin D. Heaney (OASIS, Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039), Arthur B. Baggeroer (MIT, Cambridge, MA 02139), Peter F. Worcester, Matthew A. Dzieciuch (Univ. of California at San Diego, La Jolla, CA 92093), James A. Mercer, Bruce M. Howe, and Rex K. Andrew (Univ. of Washington, Seattle, WA 98105)

During the North Pacific Acoustics Laboratory (NPAL-04) experiment broadband transmissions (50 Hz bandwidth, 75 Hz center frequency, 27.28 s period length) from a bottom-mounted source off the island of Kauai were received on a towed array (350 m depth, 200 m aperture) near the island and a large aperture vertical line array (VLA) at a range of 2469.5 km. Previous work [Heaney, J. Acoust. Soc. Am. **117**(3), 1635–1642 (2005)] has demonstrated that bottom interaction near the source is a fun-

damental process in the evolution of the wave-train. In this paper results from eight receptions at ranges from 2 to 300 km and the reception on the distant VLA is used to examine the wavefront evolution as it propagates into deep water. Results from a local geo-acoustic inversion will be presented as well as comparison of data with broadband parabolic equation simulations. The initial conclusion is that the first-order bottom bounce, off the seamount near the source, acts as an image source, contributing to a complex arrival pattern even at great distances.

9:50

2aAOb9. Evolution of second-order statistics of low-order acoustic modes. Tarun K. Chandrayadula, Kathleen E. Wage (George Mason Univ., 4400 Univ. Dr., MSN 1G5, Fairfax, VA 22030, tchandra@gmu.edu), James A. Mercer, Bruce M. Howe, Rex K. Andrew (Univ. of Washington, Seattle, WA 98195), Peter F. Worcester, and Matthew A. Dzieciuch (UCSD, La Jolla, CA 92093)

Low-mode signals measured during long range tomography experiments, such as the Acoustic Thermometry of Ocean Climate (ATOC) and the North Pacific Acoustic Laboratory experiments, have a random arrival structure due to internal waves. At megameter ranges, the narrow-band mode amplitude is predicted to be Gaussian and uncorrelated with other modes [Dozier and Tappert, J. Acoust. Soc. Am. **63**, 353–365 (1978)]. Wage et al. measured the centroid, frequency coherence, time spread, and time coherence for the broadband ATOC mode signals received at ranges exceeding 3000 km [Wage et al., J. Acoust. Soc. Am. **117**, 1565–1581 (2005)]. The 2004 Long Range Acoustic Propagation EXperiment (LOAPEX) provided an opportunity to observe how the mode statistics evolves with range. This talk investigates the mean, temporal covariance, and intermodal correlation of the low modes at ranges between 50 and 3200 km using LOAPEX data. Broadband parabolic equation simulations were performed to model internal wave effects on the low-mode signals at the various LOAPEX stations. A kurtosis measure is used to study the amount of cross-modal scattering with respect to range. Statistics of the measured and simulated mode signals are compared. [Work supported by ONR.]

10:05–10:20   Break

10:20

2aAOb10. Modal group time spreads in weakly range-dependent deep ocean. Ilya A Udovydchenkov and Michael G Brown (RSMAS, AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Modal group time spreads in deep ocean environments are investigated. The environments considered consist of a range-independent background on which a range-dependent perturbation, due for example to internal waves, is superimposed. Modal group time spreads are defined to be the temporal spread of acoustic energy with a fixed mode number across a nonvanishing frequency band. It is shown that there are three contributions to modal group time spreads that combine approximately in quadrature. These contributions are (1) the reciprocal bandwidth, (2) a deterministic dispersive contribution, and (3) a scattering-induced contribution. Contributions 2 and 3 are shown to be controlled by the waveguide invariant $\beta(m,f)$, a property of the background sound speed structure. Numerical simulations are shown to be in good agreement with theoretical predictions. [Work supported by ONR.]

10:35

2aAOb11. Long-range ocean acoustic propagation experiment (LOAPEX) acoustic modeling analysis using measured environmental data. Michael Wolfson and Frank Henyey (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698)

One objective of LOAPEX concerns the range dependence of multiple forward scattering for broadband, long-range acoustic transmissions. Using environmental data extracted during the experiment, a sound speed field is constructed, which includes sound speed fluctuations due to neu-

trally buoyant intrusions as well as internal waves. This sound speed field is used as input to a full wave simulation model. Range-dependent results from simulations regarding the coherent field and depth scattering below lower turning point caustics are presented.

10:50

2aAOb12. A comparison of experimental and theoretical forward scattering by the Kermit-Roosevelt Seamount. Arthur Baggeroer, Henrik Schmidt, Joseph SikoraIII, Hyun Joe Kim (MIT, 77 Massachusetts Ave., Cambridge, MA 02139), Edward Scheer (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Peter F. Worcester, Matthew A. Dzieciuch (Univ. of California at San Diego, La Jolla, CA 92093), and Kevin D. Heaney (OASIS Inc., Fairfax Station, VA 22039)

The SPICEX-LOAPEX-BASSEX experiments measured forward scattering from the Kermit-Roosevelt seamounts. LFM and M-sequence signals from 250-Hz moored sources and a 75-Hz ship-deployed source were received with Penn State University's 162-element FORA array. Receptions were obtained from these sources at ranges of hundreds to 1600 km. Previous work [J. Acoust. Soc. Am. **117**, (2549)] discussed the BASSEX forward shadow and patterns of horizontal refraction around the seamount. Additional work [J. Acoust. Soc. Am. **118**, (1936)] compared experimental results with theory using a conical section model of a seamount. This presentation concerns the pulse compressed and array processed time series of M-sequence signals in the forward scattered field of the seamounts. Both travel times and angles of arrival are reconciled with 2-D ray-trace models. Results are comparable to refracted rays, which mainly propagate over the seamount. These, however, do not explain additional echoes, which appear in the experiment, possibly due to reflection and/or diffraction. Normal mode and PE methods are used to confirm ray models and identify additional echoes.

11:05

2aAOb13. On the forward problem of mode coupling through shallow-water nonlinear internal waves. John Colosi and Aaron Young (Dept. of Oceanogr., Naval Posgraduate School, Monterey, CA 93943)

Using the Dyson series, we have developed a 2-D (depth and range) theory to predict the coupling of low-frequency acoustic modes propagating through shallow water nonlinear internal wave fields. Parabolic equation numerical simulations with typical shallow water internal wave fields show that the theory can accurately predict mode coupling up to an acoustic frequency of about 300 Hz. Second-order theory may extend the results to higher frequency and larger amplitude internal waves. Application of the theory to 3-D problems in which azimuthal coupling is weak will be discussed.

11:20

2aAOb14. High-frequency current tomography in coastal regions. Mohsen Badiey and Entin Karjadi (College of Marine and Earth Studies, Ocean Acoust. Lab., Univ. of Delaware, Newark, DE 19716)

In deep-water tomography refracted ray paths can provide accurate account of travel time for the inverse problem. However, in shallow coastal regions, due to the geometry of the source-receiver as well as the waveguide dimensions, most acoustic energy is a multiple reflection from sea surface and sea bottom. This makes the forward problem which deals with resolving the multipath structure of the arriving signals difficult. In this paper, we discuss the limits of reciprocal acoustic transmission as a function of waveguide temporal and spatial parameters for the purpose of

tomography application in coastal regions. Results from a high-frequency (0.5–24 kHz) quasi-reciprocal tomography experiment in a very shallow-water coastal region are presented. Using these data, the accuracy and limit of current tomography for variable sea states are examined for different center frequencies and bandwidths. [Work supported by ONR.]

11:35

2aAOb15. A multistation coastal acoustic tomography system with coherent operation. Hong Zheng (S.E.A. Corp., IU Bldg. 4F, 2-23 Shiohama, Ichikawa, Chiba 272-0127, Japan), Arata Kaneko, and Noriaki Gohda (Hiroshima Univ., Higashi-Hiroshima, Hiroshima 739-8527, Japan)

A multistation coastal ocean acoustic tomography (CAT) system with coherent operation has been developed and applied to field experiments over ten times in the last 8 years. The system is designed to measure both temperature and current velocity fields in the coastal ocean. As is well known, an ocean acoustic tomography is generally based on travel time measurements. In the CAT system, the clock accuracy for travel time measurement is most important to realize coherent transmission from multistations and to measure especially current velocity. For highly accurate coherent transmission from multistations, each station is equipped with a timing module synchronized by the clock signals of global positioning system (GPS). As a result, time error is maintained less than 0.1 μs. In the sound transmission from multistations, the code division multiple access (CDMA) and spread spectrum technology are applied. The coded signals also improved remarkably the signal-to-noise ratio of the CAT system. With detecting not only the amplitude, but also the phase of received signals, the accuracy of travel time measurement is improved within a sampling interval of data. This multistation CAT system served to map 2-D time-varying tidal current structures in Tokyo Bay (Kaneko *et al.*, 2005).

11:50

2aAOb16. Nonlinear optimization for beamforming a geometrically deficient vertical line array: Application to sediment tomography. Steven E. Crocker (Naval Undersea Warfare Ctr., Newport, RI 02841), James H. Miller, Gopu R. Potty (Univ. of Rhode Island, Narragansett, RI 02882), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The 1996 Shelfbreak PRIMER experiment included broadband acoustic transmissions in the vicinity of a shelfbreak front collected using a geometrically deficient vertical line array (VLA). The broadband signals were used to tomographically image sediment sound speed and attenuation. The VLA spanned the lower half of the water column. Hence, the orthonormal properties of the underlying normal modes cannot be guaranteed for any replica vector based on the local modal structure. This produces modal sidelobes in the output of the geometrically deficient VLA. Various methods of extracting the normal modes from VLA data have been developed including a generalized least squares method shown to improve the performance of a geometrically deficient VLA processing phase encoded tomography signals [Chiu *et al.*, IEEE J. Ocean. Eng. **22**, 522–533 (2005)]. The present study seeks to enhance the modal resolution for the reception of high-amplitude, impulsive signals transmitted over a wide variety of source receiver paths in shallow water where ensemble averaging is not an option. The method uses nonlinear optimization techniques to compute a set of array element coefficients that enhance suppression of modal sidelobes, while maintaining adequate spatial matching with the desired mode. [Work supported by Office of Naval Research.]

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Shock Wave Therapy I

Robin O. Cleveland, Cochair

Boston Univ., Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

S. H. R. Hosseni, Cochair

Tohou Univ., Inst. of Fluid Science, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

Invited Papers

8:00

2aBB1. Extracorporeal cardiac shock wave therapy ameliorates myocardial ischemia in animals and humans. Hiroaki Shimokawa (Grad. School of Medicine, Tohoku Univ., Sendai 980-8574, Japan)

The number of patients with end-stage coronary artery disease without indication of coronary intervention or coronary artery bypass surgery has been increasing. Gene and/or cell angiogenic therapies are currently under investigation for those patients, however, the therapies are invasive in nature and their effectiveness and safety have not been established yet. We here report our successful development of noninvasive extracorporeal cardiac shock wave therapy (CSWT) for the treatment of myocardial ischemia. We first confirmed that endothelial expression of VEGF and its receptor Flt-1 peaks with a low level of SW at 0.09 mJ/mm^2 *in vitro*. We then demonstrated that CSWT effectively ameliorates myocardial ischemia in a porcine model of effort angina without any adverse effects. Finally, we were able to demonstrate that CSWT effectively ameliorates myocardial ischemia and related symptoms in patients with end-stage coronary artery disease without any adverse effects. Myocardial perfusion was improved with CSWT only in the ischemic region where we treated with the therapy. We are currently performing a double-blind clinical study to further confirm the usefulness and safety of our CSWT. [Work supported in part by grants from the Japanese Ministry of Health, Labor and Welfare.]

8:20

2aBB2. Various old and new shock-wave applications require different shock-wave beam characteristics. Reiner Schultheiss (Tissue Regeneration Technologies, LLC, 110 Arnold Mill Park, Ste. 400, Woodstock, GA 30188)

Four different medical applications of shock waves and the appropriate pressure field will be discussed. Lithotripsy requires maximum shock-wave energy at the stone. High-pressure amplitudes and the total acoustical energy are the preferable parameters discussed for the disintegration effect. Improved kidney stone disintegration coincident with reduced side effect on the nephritic tissue was realized by a low-angle focusing ellipsoid (LithoGold). In the treatment of soft tissue, unfocused shock waves allow the mechanical parameters like peak pressure and the energy density to be lowered dramatically. Activation of the body's healing mechanisms are found after low pressure/ energy application by the release of growth factors, increased number of blood vessels, and the formation of new true tissue instead of scar tissue. For the treatment of diabetic skin ulcers an unfocused shock head is in controlled clinical studies within the USA. The last application relates to treating open heart muscle and requires a very small shock-wave head. A very short focal distance (11 mm) is used. New blood vessels and the formation of new heart muscle cells are found after treatment. Technical and medical reasons for the different shock-wave beam characteristics are provided.

8:40

2aBB3. Shock waves in orthopedics. Takashi Saisu (Div. of Orthopaedic Surgery, Chiba Children's Hospital, 579-1 Heta-chou, Midori-ku, Chiba 266-0007, Japan, h-saisu@dc4.so-net.ne.jp)

Extracorporeal shock wave therapy (ESWT) in the orthopedic field was first reported in 1991 by Valchanou, for the treatment of pseudoarthrosis after fracture. Since 1995, many authors have reported on ESWT for the treatment of painful disease such as calcifying tendinitis of the shoulder, tennis elbow, or planter fasciitis. Now, it has become one of the major treatment methods in orthopedics and veterinary medicine, especially for racehorses. The shock wave research group at Chiba University has been studying ESWT since 1993. We have been elucidating mechanisms of action to the bone and the neural system, performing basic studies using rabbits and rats. A shock wave has three different effects, which are bone induction after multifocal microfracture, perforation of calcified tissue, and induction of degeneration of free nerve ending. Each effect has been utilized for the treatment of pseudoarthrosis, calcifying tendinitis of the shoulder, and enthesopathy such as tennis elbow or planter fasciitis, respectively. Considering our experiences in animal studies, future applications of ESWT would be noninvasive thickening and lengthening of bone and nonsurgical acetabular plasty to prevent osteoarthritis of the hip. Recently, we have been trying nonviral gene transfer assisted by shock waves and we have succeeded in a preliminary animal study.

3064   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   3064

9:00

2aBB4. Influence of focusing and signal shape on the efficiency of stone fragmentation *in vitro* under equal energy conditions.
Rainer E. Riedlinger, Marko Liebler, and Thomas Dreyer (Univ. Karlsruhe, IHE-Akustik, Kaiserstr. 12, D-76131 Karlsruhe, Germany, rainer.riedlinger@ihe.uka.de)

To investigate the mechanisms of fragmentation in lithotripsy, simulations and experiments were performed. Wave shapes deduced from measurements near the focus of a piezoelectric source were used as input signals for finite element calculations of stresses and strains in spherical stones of different sizes. The effects of strongly and weakly focused sound fields were compared to those of a plane wave assumption representing lithotripters with large focal widths. The energies of the fields were matched. In targets with diameters less than 6 mm the simulated tensile stresses induced by strong and weak focusing exceed those of the plane wave assumption. For *in vitro* experiments a double-layered self-focusing piezoelectric source was used. To represent lithotripters with different sound fields the signal shape was varied by changing the interpulse delay but maintaining the energy. The total number of pulses needed for fragmentation into particles smaller than 2 mm diameter was counted. At usually applied energies experimental results show that the numbers of pulses needed for complete fragmentation are almost independent of the signal shape. In conclusion, the results indicate that strongly focused lithotripters are in principle more effective than those with wide focal widths in breakage of small stones.

9:20

2aBB5. Using shock waves and micro-nanostructures for gene and drug delivery.
Mark Kendall (Australian Inst. for Bioengineering and Nanotechnology (AIBN), Univ. of Queensland, Level 6, Queensland Bioscience Precinct (Bldg. 80), Brisbane, Brisbane Qd 4072 Australia)

Efficiently and safely delivering biomolecules to the skin's immunologically sensitive cells holds the promise of advancing radical new drugs, vaccines, and immunotherapies for major diseases [D. Chen *et al.*, Expert Rev Vaccines **1**(3): 89–100 (2002)]. Here I present research into using shock waves and micro-nanostructures to meet this goal. First, I analyze the physiology and immunology of the skin target and then introduce current physical approaches for targeting these cells for the immunotherapy [M. A. F. Kendall, Vaccine May 22; **24**(21): 4651–656 (2006)]. Then, I focus on biolistics—accelerating with a helium flow micro-particles coated in DNA to ballistically penetrate the human skin in order to achieve an immunological effect [M.A.F. Kendall, Shock Waves J., **12**(1), 22–30 (2002)]. I present key findings on the engineering of these devices, skin physical interactions induced by micro-particle impact and the resultant local biological responses. Finally, I introduce a micro-nanoprojection array patch technology, highlighting key advantages over biolistics, including the prevention of cell-death from shock impact with key cells.

Contributed Papers

9:40

2aBB6. Measurement of reduced stress in model kidney stones with increased rate of shock wave delivery in lithotripsy.
Teiichiro Ikeda, Michael R. Bailey, Brian MacConaghy, Lawrence A. Crum (Ctr. for Industrial and Med. Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), and Yoichiro Matsumoto (The Univ. of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan)

Slow clinical shock wave rates more effectively comminute stones. Higher rates create more cavitation bubbles along the focusing axis. Bubble clouds potentially reflect or attenuate the shock wave and also may collapse less energetically. Here, high-speed photo-elastography was used to visualize the dynamic stress distribution inside a transparent model stone. Photo-elastography records constant-stress lines, making quantification possible. PVDF sensors (4 mm diameter) measured force on the proximal face of the stones. The impulsive force of the shock wave and the cloud collapse at various clinical rates (single shocks, 1 Hz, 2 Hz, 3 Hz) in degassed and non-degassed water were calculated from the measurements. Impulse forces from the shock wave and cavitation collapse were comparable in the range $4-7\times10^{-4}$ Ns. At clinical rates in gas-saturated water, the stress fringes of the tensile component of the shock wave were reduced; the observable maximum stress was decreased; and impulsive force from the shock and the cavitation were decreased over single shocks. The results are evidence of reduced stress in the stone at higher rates due to attenuation by bubbles and less so to softened cavitation collapse. [Work supported by Grant Nos. NIH DK43881 and NSBRI SMS00402.]

9:55–10:10   Break

10:10

2aBB7. Numerical study of the collapse of a bubble subjected to a lithotripter pulse.
Eric Johnsen and Tim Colonius (Div. of Eng. and Appl. Sci., California Inst. of Technol., 1200 E. California Blvd., Pasadena, CA 91125, ejohnsen@caltech.edu)

In shock wave lithotripsy, the combined effect of focused shock waves and cavitation pulverizes kidney stones. Although cavitation has been shown to play an important role in the stone comminution process, the underlying mechanism has yet to be fully understood. The goal of the present study is to quantify the potential damage caused by the collapse of a single bubble near a solid surface. Using a high-order accurate quasi-conservative, shock- and interface-capturing scheme [E. Johnsen and T. Colonius, J. Comput. Phys., in press (2006)], the response of an air bubble subjected to a lithotripter pulse is considered. In particular, quantities important in cavitation erosion, such as wall stresses and reentrant jet velocity, are measured as functions of the properties of the pulse (amplitude, wavelength) and the geometry of the problem (stand-off distance from the wall, presence of neighboring bubbles). Preliminary two-dimensional results for a "cylindrical bubble" show that a large water-hammer pressure is measured along the wall. This pressure increases for smaller stand-off distances and longer wavelengths. Further results for spherical bubbles will be presented at the meeting. [This work was supported by NIH Grant PO1 DK043881.]

10:25

2aBB8. The effects of plane shock wave on the endothelial cells *in vitro* and the numerical analysis of pressure wave in the cell. Masaaki Tamagawa, Seiya Iwakura, Shiro Suetsugu, and Jun Fujii (Kyushu Inst. of Technol., Hibikino 2-4, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0196, Japan, tama@life.kyutech.ac.jp)

Shock wave phenomena in living tissues are being applied in the fields of tissue engineering. It is expected that the elastic cells such as endothelial cells and cardio cells have a potential to regenerate or grow quickly by using high pressure wave and short duration time from 1 μs to 100 ms. There are two mechanisms for this, one is the effect of shock waves and the other is the effect of cavitation bubbles. In this paper, to investigate only the effects of plane shock waves on the endothelial cells *in vitro*, the cells worked by shock waves are observed by microscope, and FEM numerical simulation is done to analyze the mechanism. For experiments, the shock tube apparatus is used that makes plane waves and has several tens of ms duration time. After working shock waves (peak pressure 0.4 MPa) on suspended cells, 10% of them are disintegrated, but the others increase faster than control cells *in vitro*. The growth of cells by plane shock waves is also confirmed by the fact that activated gene expression is enhanced. Numerical simulation results show the gradient of the shear stress wave is an important factor.

10:40

2aBB9. Acoustic coupling can be problematic in shock wave lithotripsy. Yuri A. Pishchalnikov, James A. McAteer, Joshua S. Neucks, R. Jason VonDerHaar, Irina V. Pishchalnikova, and James C. Williams, Jr. (Dept. of Anatomy and Cell Biol., School of Medicine, Indiana Univ., 635 Barnhill Dr., Indianapolis, IN 46202-5120, yura@anatomy.iupui.edu)

Research to improve lithotripsy has yielded considerable progress in determining mechanisms of SW-action, has led to development of innovative SW-delivery systems, and has offered practical strategies to enhance stone breakage and minimize renal injury. Little attention has been paid to how SWs are coupled to the body. *In vitro* tests were performed to assess the quality of coupling with a clinical Dornier DoLi-50 lithotripter. A test tank with a Mylar window was coupled to the treatment head using lithotripsy gel. Application of gel as performed in patient treatment always trapped air pockets at the interface, and it took considerable manipulation to eliminate them. Stone breakage was less efficient when air pockets were present, with a linear correlation between percent area of air-pocket coverage and reduction in breakage ($p < 0.001$). Breaking contact and recoupling—as can also occur in patient treatment—reduced stone breakage to ~10% of air-free coupling. Air pockets created when gel was applied to both surfaces, or with coupling and recoupling, reduced the energy density of the pulse to ~30%–50% of air-free. Thus, even under controlled conditions, acoustic coupling in SWL can be inefficient. Poor acoustic coupling could be a significant source of variability in clinical shock wave lithotripsy.

10:55

2aBB10. Experimental application of Holmium:yttrium-aluminum-garnet laser-induced shock wave as a drug delivery system. Tomohiro Ohki, S. H. R. Hosseini, Kazuyoshi Takayama (Biomed. Eng. Res. Organization, Tohoku Univ., Sendai, Japan), Junichi Sato, Atsuhiro Nakagawa, Teiji Tominaga (Tohoku Univ. Grad. School of Medicine, Sendai, Japan), Mariko Kambe (Senseki Hospital, Ishinomaki, Japan), and Mingu Sun (Tohoku Univ., Sendai, Japan)

One of the possible applications of shock waves is enhancement of drug delivery in the central nervous system. To achieve such a goal, it is necessary to develop a suitable shock wave source with the ability to be integrated with an endoscope or a catheter. We have developed a shock wave generator using a pulsed Ho:YAG laser as the energy source; this device is remarkable for its ability to expose shock wave to target tissue in a localized area. The physical feature and characteristics of the generator were clarified. The overpressures of Ho:YAG laser-induced cavitational shock waves were measured by using a PVDF needle hydrophone. Maximum overpressure of 19 MPa at 4-mm distance from tip of the device was

obtained. As a preliminary experiment, the device was applied to evaluate the enhancement of chemotherapeutic effects of shock wave *in vitro*, using human gastric cancer cell (GCIY cell) line and an anticancer drug (Bleomycin: BLM). Proliferation rate of the cells decreased to 90% and 45%, after application of 4000 and 6000 shots of shock waves at the BLM concentration of 1 mic-g/ml. Inhibitions in proliferation rate were observed in accordance with the number of shock wave applications and BLM concentrations.

11:10

2aBB11. Experimental study of micro shock wave focusing for precise medical applications. S. H. R. Hosseini, Tomohiro Ohki, and Kazuyoshi Takayama (Biomed. Eng. Res. Organization, Tohoku Univ., 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan)

For applying shock waves to sensitive medical procedures like cranioplasty in a close vicinity of the brain or treatment of myocardial dysfunction, generation of micro underwater shock waves plays an important role. Such delicate applications make limits on usage of conventional shock wave (SW) sources. In the present research a half-ellipsoidal cavity with 20.0-mm minor diameter and the ratio of major to minor diameters of 1.41 was designed as a compact extracorporeal SW source. Silver azide AgN$_3$ pellets ranging from 1.0 to 20 μg with their energy ranging from 1.9 to 38 mJ were used to generate shock waves at the first focal point $F1$ inside the reflector. Irradiation of a Q-switched Nd:YAG laser beam through a 400-μm optical fiber was used to ignite the pellet. The whole sequences of the shock wave generation, propagation, and focusing were visualized by quantitative double pulse holographic interferometry and time-resolved high speed shadowgraph methods. Pressure histories were measured at the shock wave focusing area $F2$ by using a fiber optic probe hydrophone. It is concluded that the compact ESW source has suitable characteristics for medical applications.

11:25

2aBB12. Photo-elastic, high-speed images of stress induced in cylindrical model kidney stones by lithotripsy. Brian MacConaghy, Teiichiro Ikeda, Michael R. Bailey, Adam D. Maxwell (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), and Oleg A. Sapozhnikov (Moscow State Univ., Moscow, 119992, Russia)

Recent experiments and calculations show that a focusing shear wave generated by the shock wave traveling along the length of a cylindrical stone creates the dominant stress and causes fracture of cylindrical model stones. A small disk placed on the proximal face of the stone suppressed the longitudinal wave responsible for spallation in calculations yet had little effect on the number of shock waves required to fracture the stone in experiment. However, a disk placed around the stone blocked the shock wave traveling along the stone in calculations and suppressed fracture in measurements. The conclusion was that so-called dynamic squeezing was a dominant mechanism to spallation. Here, photo-elastic, high-speed imaging was used to observe the suppression of various waves in cylindrical stones made of acrylic. Glycerol was used to avoid artifact from the curvature of the stone. Comparison was made to calculations using the elasticity equations for an isotropic medium. Agreement between measurement and calculation was excellent and supports dynamic squeezing. The results help validate the model, and the technique and modeling may help us understand where and how stress is created in other shock wave therapies. [Work supported by NIH DK43881 and NSBRI SMS00402.]

11:40

2aBB13. Shock-wave-based biolistic apparatus. Viren Menezes, S. H. R. Hosseini, Kazuyoshi Takayama (Biomed. Eng. Res. Organization, Tohoku Univ., Sendai, Japan), Mutsumi Nakada, and Akira Kanno (Tohoku Univ., Sendai, Japan)

A shock-wave-based biolistic device has been developed to deliver DNA/drug-coated micro-projectiles into soft living targets. The device consists of an Nd:yttrium aluminum garnet laser, equipped with a suitable

optical setup and a thin aluminum (Al) foil (typically 100-μm thick). The powdered vaccines to be delivered are deposited on the anterior surface of the foil and the posterior surface of the foil is ablated using the laser beam. The ablation launches a shock wave through the foil that imparts an impulse to the foil surface, due to which the deposited particles accelerate to high velocities that are sufficient to enable them to penetrate soft targets.

The device has been tested for *in vivo* DNA delivery by delivering plasmid-coated, 1-μm size, gold (Au) particles into onion and tobacco cells. The GUS activity was detected in the onion and tobacco cells after the addition of an artificial substrate. The present device is totally nonintrusive in nature and has a potential to get miniaturized to suit the existing medical procedures for drug delivery.

Session 2aEA

Engineering Acoustics: Ultrasonic Functional Devices I

Stephen C. Thompson, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Minoru Kurosawa, Cochair
Tokyo Inst. of Technology, G2-32, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan

Invited Papers

8:00

2aEA1. Ultrasonic motors—breakthrough trials for the next generation.　Yoshiro Tomikawa, Chiharu Kusakabe, Hideki Tamura, Seiji Hirose　(Dept. of Elec. Eng., Yamagata Univ., Yonezawa, Yamagata 992-8510, Japan, tomikawa@yz.yamagata-u.ac.jp), Takehiro Takano　(Tohoku Inst. of Tech., Sendai, Miyagi 982-8577, Japan), and Manabu Aoyagi　(Muroran Inst. of Tech., Muroran, Hokkaido 050-8585, Japan)

Twenty years and more have passed since real research of ultrasonic motors was started and now their practical applications are performed in some engineering fields. Therefore, such a problem has happened now as to what kinds of breakthrough ultrasonic motors must be researched for the next generation. For this reason, we are challenging the breakthrough research of the ultrasonic motors. This paper deals with such breakthrough motors. In the first part, we deal with thin-form ultrasonic motors of high revolution speed for their use as HDD motors, applying lead-free piezoelectric single crystal of LiNbO$_3$. The second part contains some gyro-moment motors, devised by us originally. The motors can be driven electromagnetically and, moreover, electrostatically in theory. Therefore the motors are suited to ones fabricated by MEMS technology and might be applied as a micro-fun motor for heat radiation in electronic apparatuses, a windmill motor in fuel cells, and so on.

8:20

2aEA2. A constant beamwidth transducer for ultrasonic applications.　Kim C. Benjamin, A. Lee Van Buren, Jeffrey A. Szelag, and Sheridan Petrie　(Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841)

A new class of transducers that produce uniform wide beams ($\sim$20 deg) at ultrasonic frequencies ($\sim$1–2 MHz) is being developed by the US Navy for ultrasonic applications. These transducers will provide a wide insonified field of view for two-dimensional ultrasonic imaging arrays. The design features a larger volume of active material, relative to conventional planar designs, which allows the application of higher drive fields. The prototype described in this article was driven to 200 VRMS and produced a sound pressure of 203 (dB *re* Pa at 1 m) at a frequency of 2.0 MHz. The details of the ultrasonic constant beamwidth transducer (CBT) design, its fabrication, and measured performance are presented. The ultrasonic CBT described in this paper will soon be incorporated into the Navy's Underwater Sound Reference Division (USRD)'s Transducer Standards Loan Program.

Contributed Papers

8:40

2aEA3. Traveling wave power in an energy circulation surface acoustic wave motor stator.　Minoru Kurosawa, Yoshito Miyazaki, and Takashi Shigematsu　(Tokyo Inst. of Technol., G2-32, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan)

Surface acoustic wave (SAW) motors have been demonstrated their high performance as micro linear motor in speed, thrust, fine stepping motion, and so on. For application of the SAW motor to some equipment, however, the power consumption will be a problem in spite of the superior performance due to the limited power resource of a system. To avoid the large power consumption, an energy circulation driving method has been proposed and tested. The circulation factor will be discussed as concerns the performance at the energy circulation electrode and the phase shift in the system. For example, in case of the 90% circulation coefficient at the circulation electrode, eight times power of the driving power circulates as a traveling wave. But the 45 deg phase shift causes a reduction of one-fourth of the traveling wave. The energy circulation of the SAW motor will be discussed quantitatively. [Research was supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Science Research.]

2a WED. AM

8:55

2aEA4. Modeling and performance evaluation of an ultrasonic pump.
Takeshi Hasegawa, Daisuke Koyama, Kentaro Nakamura, and Sadayuki Ueha (P&I Lab., Tokyo Inst. of Technol., 4259-R2-26, Nagatsuta, Midori-ku, Yokohama, 226-8503 Japan, hase@sonic.pi.titech.ac.jp)

This paper describes a low-profile miniature ultrasonic pump consisting of a disk-bending vibrator and a pipe-end located vertically to the disk center via a small gap. First, we explored the relationships between the pump performance and the geometrical conditions of the pump experimentally: the pipe diameter, the gap, and vibration distributions. Based on those results, the thickness distribution of the bending disk was designed using finite-element simulations so that the vibration distribution near the pipe might become uniform. Maximum pump pressure of 20.6 kPa was obtained with a 30-mm-diam bending disk and electrical power consumption of 3.8 W. For optimization, we constructed a model of the pump operation in this study. Nonlinearity in the elastic constant of the working fluid was introduced into the model because the temporal change rate of the gap caused by the ultrasonic vibration of the disk vibrator considerably affects the pump's pressure generation. Results show that the calculated pump pressure well simulates the measured dependence of the gap on the pressure. The acoustic pressure was measured using a fiber optic pressure probe and was compared to the calculated pressure value.

9:10

2aEA5. Methanol sensor using a shear horizontal surface acoustic wave (SH-SAW) device.
Jun Kondoh (Grad. School of Sci. and Technol., Shizuoka Univ., Johoku 3-5-1, Hamamatsu-shi, Shizuoka 432-8561, Japan), Yoshikazu Matsui (Shizuoka Univ., Hamamatsu-shi, Shizuoka 432-8561, Japan), and Showko Shiokawa (SAW&SPR-Tech. Corp., Hamamatsu-shi, Shizuoka 432-8011, Japan)

Measurements of liquid properties are important for industry, environmental issues, and human life. Because a surface acoustic wave on 36YX-LiTaO3 is a shear horizontal (SH) mode, the liquid-phase SAW sensor is realized using a crystal. In this paper, we present a SH-SAW-based methanol sensor that functions at high temperature. The SH-SAW propagation characteristics were calculated using a numerical calculation method considering temperature coefficients of the substrate and liquid. Calculation results indicate that the SH-SAW sensor can detect methanol solutions at temperature greater than 50 °C. Moreover, the calculated results suggest that the sensor sensitivity increases with increasing temperature. The methanol solutions were measured using the SH-SAW sensor. Concentration and temperature of the methanol solution were varied. Linear relationships between sensor responses and concentration were observed at each temperature. As the numerical calculation results suggest, the sensor sensitivity increases with increased temperature. Based on the experimental results, the detection limit of the sensor was estimated as 0.10% by weight at 60 °C.

9:25

2aEA6. An ultrasonic linear motor using traveling waves of twin ridge.
Osamu Koyama, Daisuke Koyama, Kentaro Nakamura, and Sadayuki Ueha (P and I Lab., Tokyo Inst. of Technol, 4259-R2-25 Nagatsutacho, Midori-ku, Yokohama-shi, Kanagawa 226-8503, Japan, shukoyama@sonic.pi.titech.ac.jp)

An ultrasonic linear motor using ridge-mode traveling waves on a substrate is proposed for precision positioning. In our experiments, two parallel ridges are machined on an alumina ceramic substrate with the spacing of 19 mm. The width, height, and length of the ridge are 3 mm, 9 mm, and 186 mm, respectively. Two piezoelectric elements, 19 mm in length and 5×5 mm^2 in cross section, were installed at the position of 15 mm from each end of the ridge, and excited at 87.136 kHz in their fundamental longitudinal mode to stimulate the fundamental ridge mode. Elliptical particle motions are formed, due to traveling waves along the ridge waveguides by the phased drive of the two transducers. An alumina ceramic slider pressed to the two waveguides is moved at 13 mm/s in the opposite direction to the traveling wave through the friction forces. The moving direction can be switched by changing the propagation direction of the traveling wave. The detailed characteristics are experimentally studied.

Invited Papers

9:40

2aEA7. An application of surface acoustic wave for micro fluid systems.
Hiroki Kuwano, Masato Sato, Jooohyng Bae, and Sumito Nagasawa (Grad. School, Tohoku Univ., 6-6-01 Aza-Aoba, Aramaki Aoba-ku, Sendai, Japan)

This paper describes an application of surface acoustic wave for micro fluid systems. Owing to its high sensitivity, fabrication easiness, and actuation capacity, current SAW devices are widely used as a chemical sensor and linear motor [Takahashi *et al.* IEEE Trans. Ultrason. Ferroelectri. Frequ. Control **43**, 901 (1996)] as well as SAW devices including filters and SAW-Tag in the field of mobile communication. Furthermore, its applications for actuators in liquid such as liquid streaming [Shiokawa *et al.*, Jpn. J. Appl. Phys. **28**(28-1), 126–128 (1989)] and atomizer [Kurosawa *et al.*, Sensors Actuators A **50**, 69 (1995)] are very attractive for micro fluid systems such as a micro injector, a micro total analysis system, and a lab-on-chip. In this paper, we propose and evaluate a novel fluid actuator device for micro fluid systems using Rayleigh waves. One of the important issues is a micro pump and fluidic friction of the micro channel for TAS. The surface acoustic wave is one solution for the issue. Some of the fluid dynamics will be discussed.

10:00–10:20   Break

10:20

2aEA8. Phased arrays and devices for high-intensity focused ultrasound.
Greg T. Clement (Harvard Med. School, 221 Longwood Ave., Boston, MA 02115)

Focused ultrasound (FUS) is rapidly attracting attention as an alternative to surgery for the treatment of tumors and other disorders. Such therapeutic procedures were first proposed over 60 years ago, but have only recently become clinically feasible. These new successes can be attributed primarily to two advances: improvement in noninvasive monitoring techniques and the development of high-power ultrasound arrays. As the field progresses, there is a demand for increased transducer bandwidth, allowing tissue-specific frequency optimization, and even imaging, while maintaining high acoustic power outputs. These arrays have been made possible via dice-and-fill manufactured 1-3 piezocomposite arrays. Transducers with over 500 elements have been tested in the clinic,

and designs have been proposed for at least 5000 elements. Meanwhile, the rapid development of high-power, broadband IC technology has made it possible to build electronic components for amplifiers and switching at a practical cost (under $50/channel). This talk will present several devices being developed in the Harvard/BWH FUS lab, in conjunction with the newly formed National Center for Image Guided Therapy. The challenges and limitations of operating at high driving intensities will also be addressed, with particular attention to nonlinear restoring behaviors of arrays driven off-resonance. [Work supported by NIH Grant U41 RR019703.]

Contributed Papers

10:40

2aEA9. Development of miniature ultrasonic probe with piezoelectric poly-crystalline film deposited by hydrothermal method. Tomohito Hasegawa, Akito Endo, Norimichi Kawashima, Shinichi Takeuchi (Toin Univ. of Yokohama, 1614 Kurogane-cho, Aoba-ku, Yokohama, Kanagawa 225-8502, Japan), Mutsuo Ishikawa, and Minoru Kurosawa (Tokyo Inst. of Technol., 4259 Nagatutamati, Midori-ku, Yokohama, Kanagawa 226-8502, Japan)

Recently, a PZT poly-crystalline film deposited on titanium substrate using the hydrothermal method was studied actively in our laboratory for development of an ultrasonic sensor. These PZT poly-crystalline films can form PZT poly-crystalline films easily on substrates of complex shape using hydrothermal method. We reported a needle-type miniature hydrophone and the ultrasonic transducer with hydrothermally synthesized PZT poly-crystalline films. Results confirmed that the hydrothermally synthesized PZT poly-crystalline films showed wideband frequency characteristics in the megahertz band of sensitivities. A miniature ultrasonic probe with wide frequency band characteristics of sensitivities is thought to be producible using hydrothermally synthesized PZT poly-crystalline films. Miniature ultrasonic probes were fabricated using hydrothermally deposited PZT poly-crystalline film on the end of 0.6-mm-diam titanium wires. Ultrasonic imaging experiments of a target in water were performed using a fabricated ultrasonic probe with hydrothermally synthesized PZT films. Consequently, it was confirmed that this probe has a center frequency of 16 MHz, a bandwidth of 80%, and distance resolution more than 80 μm. We report the structure and performance of fabricated ultrasound probes with hydrothermally synthesized PZT poly-crystalline films, along with results of imaging experiments performed in water.

10:55

2aEA10. Study of a micro-tactile sensor with longitudinal vibrator in a smart micro-ultrasonic scalpel. Shunsuke Uchikawa and Minoru K. Kurosawa (Dept. of Information Processing, Tokyo Inst. of Technol., Mailbox: G2-32, 4259 Nagatsuta, Midori-ku, Yokohama, Japan)

A smart micro-ultrasonic scalpel using a hydrothermal PZT film vibrator has been proposed and fabricated. The micro-scalpel has two functions: in addition to a micro-ultrasonic scalpel, a micro-tactile sensor to indentify tumors. The tactile sensor function is studied. The resonance frequency of the vibrator changed when the vibrator tip made contact with a rubber sample. Irrespective of compliance of the sample, the resonance frequency increased up till about tens of micrometer sample. Then, against further sample deformation by the tip, the resonance frequency decreased. From the theoretical analysis using mechanical contact modeling [K. L. Johnson, *Contact Mechanics* (Cambridge University Press, Cambridge, 1985], it was found that the resonance frequency drop was caused by increase of the damping effect of the sample due to deeper stick of the vibrator tip. The Young's modulus of samples is able to be estimated from the resonance frequency curve. From the experimental result, there was a difference in the resonance frequency between a soft rubber sample and a hard one. The elastic characteristic of the sample will be quantified.

11:10

2aEA11. The TDK model for micromachined capacitive Lamb wave transducers. Li-Feng Ge (School of Electron. Sci.e and Technol., Anhui Univ., Hefei 230039, China)

Electrostatic or capacitive ultrasonic transducers (EUTs or CUTs) were originally developed from the condenser transmitters and microphones with grooved and roughened backplates in the 1930s. Significant progress in theoretical and experimental research has been made since the introduction of micromachining techniques at the end of the 1980s. We have seen the advancement not only from the roughened ultrasonic transducers to the 2-D array micromachined capacitive ultrasonic transducers (mCUTs), but also from the grooved ultrasonic transducers to the micromachined capacitive Lamb wave transducers (mCLWTs). Thus, the theoretical TDK model originally developed for grooved ultrasonic transducers [L.-F. Ge, Proceedings 16th Intl. Conf. Acoustics/135th ASA Meeting, l08l–1082, 1998; and Sciences in China (A), (in English), **42**, 1308–1315 (1999)] can be used for mCLWTs. It is most important for a reliable model to be able to predict the transducer's natural frequencies, at which the lowest-order antisymmetric Lamb wave (A_0 wave) is excited. As an example using a typical transducer [M. H. Badi *et al.*, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**(9), 1191–1203 (2003)], the fundamental frequency predicted using the TDK model is 2.18 MHz, which agrees well with the measured result of 2.1 MHz. [Work supported by NSFC (60374044).]

Session 2aED

Education in Acoustics: Demonstrations and Tools in Acoustics Education

Uwe J. Hansen, Cochair
64 Heritage Dr., Terre Haute, IN 47803

Fumiaki Satoh, Cochair
Chiba Inst. of Technology, Dept. of Architecture and Civil Engineering, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

Chair's Introduction—9:00

After the Chair's Introduction, the authors of each paper will briefly describe their demonstrations (2 minutes each). Authors will then be at their demonstrations to interact with interested participants for the remainder of the session.

Invited Papers

2aED1. Visualization of acoustic resonance phenomena using Kundt's dust figure method. Shinichi Sakamoto, Takumi Asakura, Kanako Ueno (Inst. of Industrial Sci. Univ. of Tokyo, Komaba 4-6-1, Meguro-ku, Tokyo, 153-8505, Japan, sakamo@iis.u-tokyo.ac.jp), Yu Sakimoto, Fumiaki Satoh, and Hideki Tachibana (Chiba Inst. of Technol., Narashino-shi, Chiba, 275-0016, Japan)

It is very effective to visualize a sound field for intuitive understanding of various acoustic phenomena, especially for acoustic education. The most famous and classical visualization technique is the Kundt's dust-tube method contrived by August Adolph Kundt. He devised this experimental technique to determine the sound velocity in the air by observing the mode pattern of a standing wave excited in a glass tube. The technique can be applied to various other acoustic resonance phenomena. In Japan, Sato and Koyasu applied this technique to a two-dimensional room acoustic model experiment in which the effect of the shape of a reverberation room on the normal modes was examined. Referring to these experiments, the authors made experimental equipment to visualize acoustic resonance phenomena for an educational purpose. In our experiment, two types of two-dimensional boxes with hard surface were prepared. In these boxes, normal modes in a closed sound field and the Helmholtz resonance phenomena, which are essential and important for architectural acoustics, can be visualized. These physical experiments are visually impressive on students in architectural courses and therefore the experiment is efficiently used in architectural acoustic courses.

2aED2. Animations of different sound field decomposition methods. U. Peter Svensson and Bård Støfringsdal (Acoust. Group, Dept. of Electron. and Telecommunications, Norwegian Univ. of Sci. and Technol., NO-7491 Trondheim, Norway, svensson@iet.ntnu.no)

Many methods for solving sound-wave propagation problems are based on a decomposition of the sound field. Here, several such decomposition methods will be demonstrated and it will be illustrated how they can give the same final solution even if the partial solutions, or the solutions outside a limited domain, can be very different. Animations of harmonic and transient solutions will be used to illustrate sound-field synthesis based on a decomposition into plane waves, cylindrical harmonics, and geometrical acoustics plus edge-diffraction components. It will also be shown how a sound field that is measured in discrete points can be converted to different representations. These decompositions have pedagogical merits but can also be used in sound-field reproduction techniques. Examples will include free-field situations as well as indoor cases with a multitude of reflections. [This research was supported by the Research Council of Norway.]

2aED3. Introduction of sound material in living environment 2004 (SMILE 2004): A sound source database for educational and practical purposes. Keiji Kawai (Grad. School of Sci. and Technol., Kumamoto Univ. 2-39-1 Kurokami Kumamoto 860-8555, Japan, kawai@arch.kumamoto-u.ac.jp), Kazutoshi Fujimoto (Kyushu Univ.), Teruo Iwase (Niigata Univ.), Tetsuya Sakuma (The Univ. of Tokyo), Yoshito Hidaka (Tohwa Univ.), and Hirohito Yasuoka (The Ctr. for Better Living)

Sound sources are used in various settings of education, research, and business that are related to architectural acoustics. Especially in educational scenarios, they are very effective for students to learn theories together with the experience of listening to actual sounds related to them. Sound Material In Living Environment (SMILE 2004), a sound source database published in 2004 in DVD form (Japanese only), is the culmination of our project. SMILE 2004 is intended as a collection of sample sounds for use in lectures on architectural acoustics at universities or industrial high schools. In addition, it is useful for measurements or demonstrations for academic or practical purposes. The database contains 913 digital sound files of real sounds and impulse responses, which cover the

study fields of acoustics, noise control, and soundscapes. Each sound is accompanied by such information as the 1/1 octave band sound pressure levels and a brief explanation and photographs of the situation in which the sound was recorded. In addition, sample presentations for lectures and practical demonstrations are attached as examples of the use of this database for educational and practical purposes. Details of all the above features will be introduced in a demonstration.

2aED4. A toy using vibrations. Kentaro Nakamura and Sadayuki Ueha (Precision and Intelligence Lab., Tokyo Inst. of Tech., R2-26, 4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Tokyo, Japan)

This study is focused on a toy using vibrations. The toy is composed of a rectangular wooden bar with ten notches and a thin wooden plate at one end of the bar. The plate has a hole at its center and is loosely put at the bar end using a pin. The plate can rotate freely. If notches on the bar are scratched by another bar, the plate begins to rotate due to the vibrations excited on the rectangular bar. The interesting thing is that the rotation direction can be controlled if the way of rubbing is altered. The operation principle of the toy is revealed through experimental observations. The measurements of vibration locus of the pin help us to understand the rotation mechanism. According to these results, the method to control the rotation direction is explained and the analogy to the operation principle of ultrasonic motors is presented.

2aED5. The multimedia simulator of radiation sound field created by a vibration disc. Kohei Michigami, Takahiro Masuda, Hiroshi Suda, and Yoshiro Miida (Chiba Inst. of Technol., 2-17-1 Tsudanuma, Narashino_city, Chiba 275-0016, Japan, kohei@miida.net.it-chiba.ac.jp)

In the field of acoustic education, acoustic phenomena are represented in textbooks as sound models and complex numerical formulas. For that reason, it is difficult for students to visualize, intuitively, the relations between acoustic phenomena and complex numerical formulas. Sufficient understanding was not obtained using static teaching materials, such as sentences and figures in textbooks and notes on a blackboard. Therefore, we have produced multimedia teaching materials that render difficult phenomena as interactively visible. The multimedia teaching materials that have been produced comprise explanatory slides, relation movies, and simulator teaching materials. They show a radiation sound field created by a vibrating disc simulator. It is necessary to calculate many complex series to compute the sound field. Then, the sound field is changed complexly and remarkably by wavelength. Therefore, it is a difficult field to understand for students. Our multimedia teaching material expresses it using 3D-CG. Students can see details of the sound-field changing perspective and the wavelength and disc size ratio. Moreover, students can use it to verify the sound pressure level of a graphically selected point, just as with a measuring instrument.

2aED6. Technical listening training: Systematic training program designed to improve auditory sensitivity. Kazuhiko Kawahara, Masayuki Takada, Shin-ichiro Iwamiya, Yoshitaka Nakajima (Dept. of Acoust. Design, Faculty of Design, Kyushu Univ., 4-9-1 Shiobaru Minami-ku, Fukuoka, 815-8540, Japan, kawahara@design.kyushu-u.ac.jp), and Kazuo Ueda (Kyushu Univ., Fukuoka, 815-8540, Japan)

Technical listening training is a systematic training program designed to improve auditory sensitivity. Through this program, students can obtain the necessary auditory sensitivity before gaining work experience on the job. The type of auditory sensitivity acquired through experience depends of course on the occupation. In contrast, technical listening training provides wide varieties of experiences, so that students can obtain auditory sensitivities associated with various fields and easily adapt to new auditory environments. Technical listening training was developed at the Department of Acoustic Design of the Kyushu Institute of Design. The program has been improved and is still included as part of the curriculum of the Department of Acoustic Design. The present demonstration exhibits the concept of technical listening training as well as its practical application at the School of Design, Kyushu University.

2aED7. An integrated system for learning musical theories from western classical and popular music. Masanobu Miura (Dept. of Media Informatics, Faculty of Sci. and Technol., Ryukoku Univ., Ostu 520-2194, Japan), Norio Emura, and Masuzo Yanagida (Doshisha Univ., Japan)

Presented here is a support system for learning the theory of harmony, one of the main subjects in music education consisting of a set of forbidden progressions. By solving given tasks in the theory of harmony, students can learn forbidden progressions. Shown here is a system that generates all the allowable solutions for given bass tasks within triads, named ''BDS'' (Basse Donnée system), together with a checking system for answers given by students, named ''VISAGE'' (Validity Inspecting System for Answers of Given Exercises). VISAGE is designed to point out note allocations that violate forbidden progressions in an input answer if any. In case an answer of a given exercise does not violate forbidden progressions, the input answer is evaluated from a viewpoint of music. Also presented is ''MAESTRO'' (Musical Aesthetics Evaluation System for Tonal music in Regular and Orthodox style), which evaluates

2aED8. Acoustic demonstrations for education in speech science. Takayuki Arai (Arai Lab., Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan)

Acoustic demonstrations are highly effective for education in speech science. We recently developed two educational tools for acoustics. The first educational tool is a set of physical models of the human vocal tract and related models. It contains cylinder and plate-type models [T. Arai, J. Phonetic Soc. Jpn., **5**(2), 31–38 (2001)], a sliding three-tube model [T. Arai, Acoust. Sci. Technol., to be published], lung models, an artificial larynx, and head-shaped models [T. Arai, Acoust. Sci. Technol., **27**(2), 111–113 (2006)]. Each model has its own advantages and, if combined effectively, can produce a systematic and comprehensive education in speech production from the lungs to the head. The second educational tool is "Digital Pattern Playback (DPP)" [T. Arai *et al.*, Acoust. Sci. Tech., to be published], which converts a spectrographic image of a speech signal back to sound by digital signal processing. A printed spectrogram on a sheet of paper can also be converted immediately after capturing the image from a camera, and we confirmed that this is more intuitive for learners than converting from an electronic image. [Work partially supported by JSPS.KAKENHI (17500603).]

2aED9. Tools for speech perception, production, and training studies: Web-based second language training system, and a speech resynthesis system. Reiko Akahane-Yamada, Takahiro Adachi (ATR, 2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan), and Hideki Kawahara (Faculty of Systems Eng., Wakayama Univ. 930 Sakaedani, Wakayama, 640-8510 Japan)

We describe a Web-based instruction system for second language (L2) learners, dubbed ATR CALL. This system is a collection of speech perception, production, and comprehension training tools, designed to examine L2 acquisition mechanisms, especially for native speakers of Japanese who are learning English as L2. Each training component focuses on the acoustic-phonetic, prosodic, lexical, or semantic decoding level of spoken language. The stimuli are selected and presented from a large speech database of 13 000 English words and 6000 English sentences. In the pronunciation training component, learners' productions are scored using HMM-based evaluation algorithms, and automatic and immediate feedback about pronunciation goodness is provided. Speech analysis tools for language learners are also provided. We also describe a speech manipulation system, dubbed STRAIGHT [Kawahara *et al.*, Speech Commun. (1999)]. This system resynthesizes stimuli with quality as high as the original stimulus even after manipulating multiple parameters. A speech morphing algorithm on this system enables us to manipulate various features independently [Kawahara *et al.*, ICCASP 2003]. Thus, it generates stimulus continuum varying in supra-segmental domain and/or segmental domain. [Work supported by JSPS, and e-Society by MEXT.]

2aED10. Vowel synthesis tool based on transmission line model as a web application. Kohichi Ogata, Takuro Masuya, and Yuichiro Shin (Grad. School of Sci. and Technol., Kumamoto Univ., 2-39-1 Kurokami Kumamoto 860-8555, Japan, ogata@cs.kumamoto-u.ac.jp)

This paper introduces a web application for vowel synthesis in acoustics education. Synthesis systems that provide users with interactive operation for determining vocal tract configuration and vocal fold control parameters can be effective tools for intuitive understanding of vowel production for students. For example, differences in the size of the vocal tract and vocal folds between male and female speakers can be recognized as differences in formant and pitch frequencies in speech sounds. The realization of such a system as a web application is suitable for platform independent use, considering the recent progress in information and communication technology. In our developed system, interactive operation for determining vocal tract configuration and vocal fold control parameters is achieved on a web browser by using a Java applet. These articulatory parameters are transferred to a server computer via a client/server connection using socket programming, and synthetic speech is produced using calculation programs on the server computer. The speech synthesizer installed on the server is based on the hybrid speech production model proposed by Sondhi and Schroeter. A preliminary user test showed that the system demonstrated a certain applicability as a teaching tool.

2aED11. WWW-based multilingualized system for technical listening training. Akira Nishimura (Dept. of Media and Cultural Studies, Tokyo Univ. of Information Sci., 1200-2, Yato-cho, Wakaba-ku, Chiba, Japan)

A WWW-based system for technical listening training is proposed. Messages from the system shown in a WWW browser are multilingualized using the GNU gettext mechanism. A system administrator can easily set up the system in a user's language by translating the message catalog. The WWW-based training system is superior to previous training systems in the following respects: (1) software and hardware can be organized more easily; (2) modification and creation of the contents of training are simple; (3) individual training parameters can be set for users; (4) many users can use the system individually or simultaneously; (5) the results of training for users are saved and summed individually; and (6) difficulty of discrimination training is automatically set up individually. Training experiments of intensity discrimination and frequency discrimination were performed using a prototype of the training system for subjects who were naive to auditory experiments. Significant improvements in the abilities of both intensity and frequency

discrimination were observed from results of these experiments. The system is guaranteed to be legal for use, distribution, and modification under the terms of the GNU General Public License. Demonstrations and software packages of the system are distributed at http://adlib.rsch.tuis.ac.jp/~akira/atsdemo/index_en.html.

2aED12. Wonders in perception and manipulation of speech. Makio Kashino (NTT Commun. Sci. Labs./ERATO Shimojo Project, JST/Tokyo Inst. Tech., Atsugi, Kanagawa 243-0198, Japan, kashino@avg.brl.ntt.co.jp), Hideki Kawahara (Wakayama Univ., Wakayama, Wakayama 640-8510, Japan), and Hiroshi Riquimaroux (Doshisha Univ., Kyotanabe, Kyoto 610-0321, Japan)

Under certain conditions, the auditory system can display a number of amazing behaviors. Some of these can be quite unique and may have significant importance for our understanding of the perception of sound. This presentation focuses on behaviors closely related to speech perception. One phenomenon that will be described is termed phonemic restoration. In it, sounds actually missing from a speech signal can be synthesized by the auditory system and clearly heard. Another phenomenon involves noise vocoded speech. In it, speech remains highly intelligible despite the fact that much of its spectral detail is absent. Audio demonstrations will be provided so that the phenomena can be directly experienced. Studying these phenomena will help us understand the function of the central auditory system and develop improved techniques for speech recognition and speech synthesis. This presentation will also contain a demonstration of STRAIGHT, a versatile tool allowing manipulation of various aspects of the speech signal. These demonstrations may be of interest not only to students, but also to teachers in physiology, psychology, and engineering.

2aED13. Amazing features of the auditory system. Kaoru Ashihara (AIST Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 3058566, Japan, ashihara-k@aist.go.jp), Yukio Iwaya (Tohoku Univ., Aoba-ku, Sendai 9808577, Japan), Kengo Ohgushi (Oheyama-cho Goryo, Nishikyo-ku, Kyoto, 6101102 Japan), and Yoshitaka Nakajima (Kyushu Univ., Minami-ku, Fukuoka 8158540, Japan)

Some audio and audiovisual demonstrations will be presented in order to review the unique nature of the auditory system. The demonstrations cover a wide range of topics: the gap transfer illusion, combination tone, interaction between auditory and visual information, and binaural recording and reproduction. The phenomena to be presented are closely related to the key features of hearing, such as the auditory scene analysis, speech perception, perception of the spatial information, and interaction with other modalities. Because these demonstrations are not only attractive but also of some scientific importance, they can be used as teaching materials in education of physics, physiology, psychology, and engineering. Studying these phenomena will give us insights into the functions of the peripheral and central auditory system.

2aED14. Inexpensive demonstration experiments in acoustics. Andrew Morrison (Dept. of Phys., Illinois Wesleyan Univ., P.O. Box 2900, Bloomington, IL 61702-2900) and Thomas Rossing (Stanford Univ., Stanford CA 94305)

We describe and demonstrate simple classroom demonstration experiments in acoustics using inexpensive equipment.

2aED15. Demonstration of wave synthesis and spectral analysis. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809)

Many inexpensive keyboard synthesizers include a flute sound with overtone levels about 30 dB below the fundamental. Since these wave shapes are nearly perfect sine waves, they are ideally suited to introduce the concept of complex tone synthesis by adding harmonics. Such a keyboard will be available at the station for demonstration purposes and for viewer experimentation. Complex tone synthesis using amplitude control in a Cool-edit program will also be available.

2aED16. Animations, visualizations, and auralizations in architectural acoustics. Ralph T. Muehleisen (Civil and Architectural Eng., Illinois Inst. of Technol., Chicago, IL 60616, muehleisen@iit.edu)

A number of animations, visualizations, and auralizations developed for use in architectural acoustics classes are presented. Animations of plane waves, spherical waves, and standing waves were created using MATLAB software. Visualizations of modes in regular and arbitrary shaped enclosures were created in Comsol Multiphysics software. Auralizations of tonal and broadband noise, hearing loss, transmission loss, and classroom acoustics were created using Adobe Audition software.

Contributed Papers

2aED17. Sonification of sound: Auditory display for acoustics education. Densil Cabrera, Sam Ferguson, and Robert Maria (Faculty of Architecture, Univ. of Sydney, NSW 2006, Australia)

Sonification is concerned with data representation to the ear using nonspeech audio. While auralization is routinely used in architectural acoustics, the broader processes encompassed by sonification are not so commonly applied, but can enhance the appreciation of acoustical proper-

ties of materials and structures, audio signals, and acoustic system measurements. We present examples of sonifications such as (*i*) auditory graphs of reflection and transmission coefficients and reverberation times; (*ii*) auditory graphs of percentiles and moments, applied to time-varying sound levels and to magnitude spectra; (*iii*) conditioning of room impulse responses for sonification using techniques such as time stretching, signal reversal, filtering and auto-convolution/correlation; (*iv*) sonification of complex frequency response or signal spectra; and (*v*) applications of the Hilbert transform for audio signal sonification. Compared to graphic or

numeric representation, such sonifications have the advantage of providing an aural experience of the aspect of sound that is sonified, and so directly convey to a listener something of the meaning of the data.

Furthermore, since it was possible to use it for a hand bell and a musical game mimicking the card game "Concentration," it may be applicable for musical education.

2aED18. Rainfall estimation using SORFED (SOund Recording For EDucation). Lucas B. Lezamiz (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Box 355640, Seattle, WA 98105, lucas@apl.washington.edu)

Underwater sound provides an abundance of information about the environment, both natural and man-made. Using an underwater hydrophone and a webcam above the water's surface, the SORFED website gathers and stores ambient noise from Lake Union 24 hours a day, 30 days at a time. Users can explore the sound signatures of different weather events and human activity on the lake by listening to audio clips side by side with video from the webcam. This website was used as the primary tool for teaching high school students in Seattle how to estimate rainfall by listening to underwater noise. Students worked in teams to calibrate underwater noise to rain rate, then used this relationship to estimate rain rate as a function of time on other days based on the noise level recorded in the lake. Total rainfall was then found by integration. Using similar techniques, students can explore other topics including shipping rate estimation, human behavior and activity, signal-to-noise ratio, and concepts in differentiation and integration.

2aED19. Beamforming with multiple acoustic sources in the classroom. Edward Tucholski (Phys. Dept., Chauvenet Hall, U.S. Naval Acad., 572C Holloway Rd., Annapolis, MD 21402, ejtuchol@usna.edu)

Superposition of multiple sources, constructive and destructive interference, and beamforming are common topics in undergraduate acoustics. While the mathematical models of these topics are fairly easily performed by students, particularly in the far field, the learning experience is enhanced with simple visualizations and by actually building the model with speakers and measuring the resulting amplitude at various angles with a microphone. While this experiment is often performed in an anechoic chamber for best agreement between theory and experiment, the apparatus will work in a typical classroom as a demonstration. The uncomfortable tonal noise in the classroom is eliminated if ultrasonic (30–50-kHz) sources are used. Apparatuses to be demonstrated have been used successfully at the U.S. Naval Academy in both majors and basic physics classrooms, where the professional development of every officer requires an understanding of sensor beamforming.

2aED20. A MIDI (musical instrument digital interface) controller using a mat switch operable by a user in a wheelchair. Takuya Niikawa (Dept. of Biomed. Eng., Osaka Electro-Commun. Univ., 1130-70 Kiyotaki Shijo-Nawate, Osaka 575-0063, Japan), Ryosuke Kawachi, Kotaro Minato (Nara Inst. of Sci. and Technol., Nara 630-0192, Japan), Tatsuo Yoshihara, and Naoya Terayama (Osaka Electro-Commun. Univ., Osaka 575-0063, Japan)

In this study, we devised a MIDI controller that a user with severe physical dysfunction was able to operate by depressing a mat switch with an electric wheelchair, and proposed a new method for musical education and therapy employing the system. This system comprises a mat switch resistant to 100 N of a movement load in a 500×700 mm^2 area, an A/D converter, a PC, and a MIDI device. Signal output from the mat switch is converted to MIDI signals by the PC and transmitted to the MIDI device. By pushing down ten mat switches arranged in a matrix pattern, a user can manipulate all elements necessary to control MIDI signals, such as volume and musical sounds. As a result, it was confirmed that a user in an electric wheelchair was able to control musical notes, volume, tunes, and scales.

2aED21. An online virtual acoustic guitar laboratory. Nelson Lee and Julius O. SmithIII (Ctr. for Comput. Res. in Music and Acoust. (CCRMA), The Knoll, 660 Lomita, Stanford, CA 94305-8180, nalee@stanford.edu)

We present a "virtual acoustic guitar laboratory" in which students can bring in their own acoustic guitars, record several notes, and follow a step-by-step procedure to create their own real-time computational model, suitable for use in a synth plugin in an electronic music studio. The guitar synthesis model is based on the commuted waveguide synthesis technique (http://ccrma.stanford.edu/~jos/cs.html). Online laboratory materials are presently available at (http://ccrma.stanford.edu/~jos), the project website (http://ccrma.stanford.edu/realsimple/), and are also being ported to Connexions (http://www.cnx.org/), an online repository of educational materials freely distributable under a Creative Commons license. Connexion's "modules" and "courses" (which are simply module sequences) are readily and freely accessible and modifiable to allow instructors to adapt the materials as appropriate for their own classroom needs. The online lab materials contain Web pointers to relevant portions of online textbooks covering theoretical aspects of the analysis and synthesis methods used, thereby offering the student a demand-driven deeper understanding within a motivating context. [This work was supported by the Wallenberg Global Learning Network.]

2aED22. Application of speech recognition to teaching English as a foreign language. Hiroyuki Obari (Aoyama Gakuin Univ., 4-4-25, Shibuya-ku, Shibuya-ku, Tokyo, 150-8366 Japan), Hiroaki Kojima, Hiroshi Oomura (Adv. Industrial Sci. and Technol., Tsukuba, Ibaraki, 305-8568 Japan), and Machi Okumura (PRONTEST, Tsukuba, Ibaraki, 305-8568 Japan)

English is considered to be an important common international language; therefore, in Japan English education plays an important role to make Japanese speakers of English intelligible enough to be properly understood for international communication. Intonation and the rhythm of the language and pronunciation of each word are crucial to comprehensive speech. So far the CALL system has not been well equipped with effective software for improving pronunciation. This is why PRONTEST and AIST (National Institute of Advanced Industrial Sciences and Technology) have been jointly developing special software for teaching English pronunciation using the technology of Speech Recognition System integrated into the CALL system. This software can automatically correct the crucial pronunciation of lower Japanese EFL learners and teach them how to improve its pronunciation with a good piece of advice after they pronounce several words with the computer. So far we have successfully developed the CALL system, in which we can deal with the following sounds: [r] [l] [s] [sh] [th] [z] [f] [v] [d] [b]. We carried out an experiment and found out that about 90% of the students were satisfied with correcting their pronunciation using this software at Aoyama Gakuin University.

2aED23. Manuscript preparation with LyXforJASA. John C. Burgess (Dept. of Mech. Eng., Univ. of Hawaii, 2540 Dole St., Honolulu, HI 96822)

LyXforJASA is a graphical software package for creating manuscripts that meet the submission requirements of the Peer-X-Press (PXP) system used by the *Journal of the Acoustical Society of America* (JASA). Its main features are graphical templates for superscript and author-year citation styles and for single figures. It is based on LyX, a screen-based graphical user interface (GUI) to the LaTeX typesetting language. Authors see manuscript content, not LaTeX code. LyX liberates authors from document bookkeeping, such as spacing, font selection, and section, equation,

figure, table, and citation numbering. LyX has a comprehensive, point and click, GUI mathematics editor with an extensive list of symbols. It can export text in LaTeX, PDF, PostScript, and ASCII formats. It has been ported successfully to Linux, Macintosh OSX, and Windows XP. LyXfor-JASA packages can be found at http://ww2.hawaii.edu/~jcb/LyXforJASA/LyXforJASA.html

WEDNESDAY MORNING, 29 NOVEMBER 2006

MONARCH ROOM (ROYAL HAWAIIAN), 8:20 TO 11:35 A.M.

Session 2aMU

Musical Acoustics: Comparing Asian and Western Instruments I

Tomoyasu Taguti, Cochair
Konan Univ., 8-9-1 Okamoto, Higashi-ku, Kobe 658-8501, Japan

Thomas D. Rossing, Cochair
Stanford Univ., CCRMA, Dept. of Music, Stanford, CA 94305

Invited Papers

8:20

2aMU1. Acoustics of Eastern and Western bells, gongs, and drums. Thomas Rossing (CCRMA Dept. of Music, Stanford Univ., Stanford, CA 94305)

Bells, gongs, and drums have been part of most Eastern and Western cultures throughout history. We compare the acoustics of some examples of these types of percussion instruments, both contemporary and historical.

8:40

2aMU2. Acoustics of the Korean percussion instruments pyeongyeong and pyeonjong. Junehee Yoo (Dept. of Phys. Education, Seoul Natl. Univ., Seoul 151-742, Republic of Korea, yoo@snu.ac.kr) and Thomas D. Rossing (Stanford Univ., Stanford, CA 94305)

Pyeongyeong is a set of 16 L-shaped chime stones and pyeonjong is a set of 16 oval chime bells. Vibrational mode shapes and mode frequencies of the pyeongyeong determined by holographic interferometry and by experimental modal testing agree very well with those calculated using finite-element methods. A model of the pyeongyeong in which the geometry can be varied shows that the actual shape is probably the one that gave the preferred sound. Vibrational mode shapes and frequencies of the pyeonjong, also determined by holographic interferometry and by experimental modal testing, are split into doublets due to the oval shape of the bells. The splitting is less than in Chinese two-tone bells but it gives rise to audible beats, as in large Korean temple bells.

9:00

2aMU3. Acoustics of the Tibetan long trumpet. Murray Campbell (School of Phys., Univ. of Edinburgh, Edinburgh EH9 3JZ, UK, d.m.campbell@ed.ac.uk) and Arnold Myers (Univ. of Edinburgh, Edinburgh EH8 9AG, UK)

The Tibetan long trumpet, or dung-chen, is a metal horn made in several telescoping conical sections. Dung-chen range in length from around 1.3 to over 3 m; the integral mouthpiece is wide and very shallow. Identical pairs of dung-chen are used in rituals of Tibetan Buddhism, often in combination with other instruments. This paper describes bore profile and input impedance measurements on a number of examples of the dung-chen, and compares the acoustical and musical properties of these Asian brass instruments with those of Western European long trumpets and horns.

9:20

2aMU4. Sound production in Asian and Western free-reed instruments. James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402)

There are two families of free-reed instruments. This paper summarizes some of the research on both done in recent years, with emphasis on differences in history and means of sound production. The Asian mouth-blown instruments are of ancient origin and use a free reed coupled to a pipe resonator. The sheng, sho, khaen, and bawu are typical examples. These reeds are approximately symmetric, typically cut from a single strip of material, and often operate on both directions of air flow. They behave as blown-open or outward-striking reeds, with playing frequency above both the resonant frequency of the pipe and the natural frequency of the reed. The Western free-reed family originated in Europe about 200 years ago and includes the reed organ, harmonium, harmonica, and the accordion-concertina family. In these the reed tongue is offset from the opening in the frame. This permits operation on only one direction of air flow, and pipe resonators are not required and generally not used. The playing frequency of the air-driven reeds is normally below the natural reed frequency and, if the reed is coupled with a pipe, below the resonant frequency of the pipe as well.

2a WED. AM

2aMU5. Comparing the Hawaiian nose flute with the western concert flute. Lydia Ayers and Andrew Horner (Hong Kong Univ. of Sci. and Technol., Clear Water Bay, Kowloon, Hong Kong)

This presentation will compare the spectral and expressive characteristics of the Hawaiian nose flute with the western concert flute. The Hawaiian nose flute is a bamboo courtship flute. This presentation will give live demonstrations on several nose flutes and the western flute, along with brief descriptions of the musical and spectral characteristics of these instruments, and then describe the additive synthesis design template used to capture their subtle timbral characteristics using score data in Csound. In addition, the expressive design successfully uses spectral interpolation for the idiomatic ornamental phrasing for each instrument. A special amplitude/frequency function modulation method simulates the trills, which are also slurred within phrases of varying numbers of notes. [This work was supported by the RGC Competitive Earmarked Research Grant 613505, with special thanks to Rodney ''Kala'' Willis for making the nose flutes.]

10:00−10:15   Break

10:15

2aMU6. Cross-cultural comparison of flute acoustics and playing techniques. Peter L. Hoekje (Dept. of Phys. and Astron., Baldwin-Wallace College, Berea, OH 44017, phoekje@bw.edu) and Pedro Eustache (Castaic, CA 91384)

Examples of flutes from different cultures around the world will be demonstrated, including European flutes, Persian ney, Egyptian kawala, traditional Indian bansuri, custom King bansuri, Japanese shakuhachi, Chinese di, South American kena, and Polynesian nose flutes. The differences among these may in large part be classified according to the details of their sound production mechanism at the mouthpiece, and of the bore and toneholes that form the resonator. These features are closely linked with the sound of the instrument as well as the notes and music played on it, and the playing techniques and styles developed for it. A cross-cultural problem arises when a musician needs to play with the sound and style of a non-Western flute tradition in the context of Western instruments and music. Various solutions to this problem, involving adaptation of either the mouthpiece or the resonator, have been developed.

Contributed Papers

10:35

2aMU7. Tunings of historical Korean pyeongyeong and pyeonjong. Junehee Yoo, Suyoung Jin, Hyunsoo Kim (Dept. of Phys. Education, Seoul Natl. Univ., Seoul 151-742, Republic of Korea, yoo@snu.ac.kr), Koeng-mo Sung, Junguk Noh, and Hyun-woo Ko (Seoul Natl. Univ., Seoul 151-742, Republic of Korea)

In order to reconstruct the *whangjongeum* or scale in Korean traditional instruments, we have measured frequencies of historical pyeongyeong stones and pyeonjong bells mainly from the 14th to 19th centuries. Pyeongyeong consists of 16 chime stones tuned to a scale quite similar to the Western scale beginning with C4. Pyeongjong consists of 16 chime bells, tuned approximately an octave lower than pyeongyeong. We present frequency averages and standard deviations for 124 chime stones and 64 chime bells available in the National Center for Korean Traditional Performing Arts. The spread in frequencies appears to be greater in some periods in history than in others, but the average frequency is quite consistent.

10:50

2aMU8. Input impedance of asian free-reed mouth organs. Eric A. Dieckman (Truman State Univ., Kirksville, MO 63501) and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

The khaen, naw, bawu, and sheng are examples of Asian free-reed mouth organs, which incorporate free reeds coupled to pipe resonators. Previous research has shown that the reeds in these instruments behave as blown-open reeds: the playing frequency is above both the natural frequency of the reed and the first peak of the measured impedance curve. Detailed calculations of input impedance have been made for a variety of these instruments, taking into account the position of the reed along the pipe, tuning slots, finger holes, and noncircular cross sections. The details of these calculations are in good agreement with the measured impedances of the same instruments, even in the higher harmonics. A study of the coupling of the free reed and pipe resonator has provided additional details about the influence of the pipe impedance on the playing frequency and sound spectrum of the reed-pipe combination. [Research supported by NSF REU Grant PHY-0354058.]

11:05−11:35
Mini-concert of flute music by Pedro Eustache

Session 2aPA

Physical Acoustics and Noise: Sonic Boom From Supersonic Aircraft and Projectiles I

Victor W. Sparrow, Cochair

Pennsylvania State Univ., Graduate Program in Acoustics, 316B Leonhard Bldg., University Park, PA 16802

Kazuyoshi Takayama, Cochair

Tohoku Univ., Biomedical Engineering Research Organization, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

Chair's Introduction—8:00

Invited Papers

8:05

2aPA1. Model experiments to study sonic boom propagation through turbulence.　Philippe Blanc-Benon and Sebastien Ollivier (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France)

The aim of this paper is to present data from laboratory-scale experimental simulations of sonic boom propagation through a turbulent atmosphere. The source used to simulate sonic boom N waves is an electric spark source. In order to model turbulence, two setups have been used: a turbulent jet for kinematic turbulence and a heated grid for thermal turbulence. Acoustic measurements have been conducted for three configurations: free field; close to a plane surface in order to simulate ground effects; and close to a curved surface in order to investigate the propagation into a shadow zone. The surfaces were either smooth or rough. With turbulence, many waveforms have been recorded, then the statistics of various parameters, including the peak pressure and the rise time, have been analyzed. In the free field, data show that the increase of the mean rise time and the decrease of the mean peak pressure are linked to the probability of random focusing. The experiments have enabled separate assessment of the effects of turbulence and roughness in the shadow zone caused by the curved surface. Comparison with numerical simulations will be discussed. [This work was supported by European Commission Contract N:G4RD-CT-2000-00398 and French Ministere de la Recherche (Decision 00T0116).]

8:25

2aPA2. Sonic-boom research activities in the Japan Aerospace Exploration Agency (JAXA).　Yoshikazu Makino　(JAXA, 7-44-1 Jindaijihigashi-machi, Chofu, Tokyo 182-8522, Japan, makino.yoshikazu@jaxa.jp)

Some sonic-boom research activities conducted in the Japan Aerospace Exploration Agency (JAXA) are summarized. JAXA started a scaled supersonic experimental airplane program named NEXST-1 in 1996 in order to establish advanced CFD-based design technologies. The flight test of the NEXST-1 was conducted in October, 2005 at Woomera test range in Australia. The sonic-boom signatures of the NEXST-1 were measured on the ground in order to confirm the possibility of demostrating the low sonic-boom design technology with scaled experimental airplanes for a future low-boom demonstrator program. The measured sonic-boom signatures were compared to the signatures extrapolated from the near-field pressure signatures predicted by CFD and panel method or measured in the supersonic wind tunnel. Low-sonic-boom design technology using gradient-based optimization tool with CFD is developed and applied to a non-axisymmetrical fuselage design concept in order to reduce the drag penalty. JAXA works on a human acceptability study in cooperation with Kawasaki Heavy Industries (KHI) using their sonic-boom simulator. Some low-boom signatures in addition to some typical N-shaped signatures with various peak pressure levels were simulated in the boom box and used for subjective test in order to collect the human acceptability data for future sonic-boom acceptability criteria.

8:45

2aPA3. Prediction and measurement of a weak sonic boom from an entry vehicle.　Kenneth J. Plotkin　(Wyle Labs., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202, kenneth.plotkin@wylelabs.com), Russell J. Franz, and Edward A. Haering, Jr. (NASA Dryden Flight Res. Ctr., Edwards, CA 93523)

There is a current interest in measuring low-amplitude sonic booms in the atmosphere. This interest is related to verifying the predicted loudness and structure of the shock waves. The reentry of the Stardust comet sample probe on 15 January 2006 provided an opportunity for such a measurement. PCBoom4 was used, together with Stardust's projected reentry trajectory, to predict the sonic boom footprint. A theory for the sonic boom of a drag-dominated blunt hypersonic vehicle had been developed by Tiegermann [Ph.D. thesis, Cornell University, 1975]. This theory provides an effective F-function based on energy dissipation, using a blast wave analogy and numeric solution. The inputs to that model are the vehicle drag and flight conditions. Tiegermann's model was implemented in PCBoom4. Drag was obtained by computing deceleration from the expected entry trajectory and multiplying by vehicle weight. A boom with peak overpressure of 0.0524 psf, arriving at 10:00:28.465 UTC, was predicted for a planned measurement site in Ruby Valley, NV, using an atmospheric profile from the upper air reporting station at nearby Elko, NV. The measured sonic boom had a peak overpressure of 0.052 psf and arrived at 10:00:29.719 UTC.

2aPA4. Toward silent supersonic transport—A fundamental study of supersonic biplane. Kazuhiro Kusunose (Tech. Res. and Development Inst., Japan Defense Agency, Sakae-cho 1-2-10, Tachikawa, Japan 190-8533), Daigo Maruyama, Kisa Matsushima, Kazuhiro Nakahashi (Tohoku Univ., Sendai 980-8579, Japan), Hiroshi Yamashita, Masahito Yonezawa, and Shigeru Obayashi (Tohoku Univ., Sendai, Japan 980-8577)

In this research, aerodynamic design of biplane airfoils in supersonic flight is discussed based on computational fluid dynamics (CFD). In supersonic flight, airfoils generate strong sonic booms and wave drags accompanied by shock waves. New airfoil geometries which significantly reduce shock waves using a biplane concept will be proposed. The background of this concept originates from Busemann biplane and Licher type biplane concepts. In order to focus on the shock-wave characteristics around biplane configuration, inviscid flow (Euler) analyses are performed (which are particularly suitable for wave drag analyses). For the evaluation of the reduction level of shock waves, the wave drag coefficient is used. The design Mach number is 1.7. The aerodynamic design is conducted using an iterative inverse design method that is newly implemented. A biplane configuration with a desired performance has been obtained. Having total maximum thickness ratio of 0.10, it has the lift-to-wave-drag ratio of 21.7 at a desired lift condition for supersonic flight. At the range of lift coefficient more than 0.14 this designed biplane has lower wave drag than that of a (zero-thickness) single flat plate airfoil. This result will lead to the realization of silent supersonic transport.

2aPA5. Ballistic range simulation of sonic boom generation. Kazuyoshi Takayama (Biomed. Eng. Res. Organization, Tohoku Univ., Sendai, 980-8577, Japan), Ming-Yu Sun, and Takamasa Kikuchi (Tohoku Univ., Sendai, 980-8579, Japan)

A result of recent ballistic range experiments applied to a sonic boom simulation is reported. For flight simulations at transonic speed ranging from 340 to 400 m/s, we used a 51-mm bore gas gun launching 51-mm-diam sabot and projectile combinations, and for supersonic flight simulations exceeding 400 m/s we used a 15-mm bore two-stage light gas gun. Flow diagnostics were performed mainly by double-exposure holographic interferometry, shadowgraph recorded with a high-speed video camera, and pressure measurements. The present experimental setup is so versatile that we placed a water basin on the recovery tank floor and measured interactions between sonic booms and wavelets in the water basin. As one of the basic research topics related between shock dynamics and aeronautics, we launched 40-mm-diam spheres at high transonic speed of shock Mach number ranging from 1.01 to 1.i at increments of every 0.01, visualized quantitatively flows over spheres, and could determine shock standoff distances over spheres. The speed range we simulated is in the region hardly achievable in existing wind tunnels.

2aPA6. Atmospheric turbulence filter functions derived from high-fidelity measurements. Lance Locey and Victor Sparrow (The Pennyslvania State Univ. Grad. Program in Acoust., State College, PA 16803)

Efforts have been underway to develop filter functions suitable for adding turbulent atmospheric effects to theoretical low-boom waveforms. Filter functions have been created based on a measured input-output relationship. The input is a relatively clean sonic boom waveform measured at altitude by a glider, and the outputs are turbulized' waveforms measured on the ground. One input waveform and multiple output waveforms are used to represent multiple realizations of the atmosphere. Work presented in 2005 [Locey and Sparrow, *Innovations in Nonlinear Acoustics, 17th International Symposium on Nonlinear Acoustics* (American Institute of Physics, Melville, NY, 2006)] yielded an initial set of filter functions using one particular algorithm and data collected during the Shaped Sonic Boom Experiment (SSBE) in January of 2004. In this talk new results will be presented based on high-fidelity measurements made at NASA Dryden Flight Research Center in June of 2006 [T. Gabrielson *et al.*, Proc. Internoise (2006)]. Time permitting, additional methods for obtaining filter functions will be discussed, including the use of existing sonic boom propagation codes modified to include atmospheric turbulence. [Work supported by the FAA/NASA/Transport-Canada PARTNER Center of Excellence for Aircraft Noise and Aviation Emissions Mitigation.]

2aPA7. Atmospheric effects on sonic boom loudness. Anthony R. Pilon (Adv. Development Programs, Lockheed Martin Aeronautics Co., 1011 Lockheed Way, Palmdale, CA 93599, tony.pilon@lmco.com)

The effects of sonic boom propagation through nonstandard atmospheres have been studied. Variations in atmospheric temperature, pressure, humidity, and winds can have significant impact on the structure of the shock waves in the boom signal, as well as the propagation path of the boom. In the study, a hybrid time-frequency domain code was used to calculate finely resolved sonic boom signals for a variety of atmospheric conditions. Idealized near–field signals for low amplitude "N-wave" and "shaped" booms were used to calculate ground level signals. The perceived level metric of Stevens ["Perceived level of noise by Mark VII and decibels (E)," J. Acoust. Soc. Am. **51**, 575–601 (1972)] was used to quantify the loudness of the calculated booms. Results of the study will be used to determine the impact of atmospheric variations on the acceptability of sonic booms from future supersonic cruise aircraft.

2aPA8. Discrete-time modeling of atmospheric turbulence with application to sonic booms. Joe Salamone (Gulfstream Aerosp. Corp., P.O. Box 2206 M/S R–01, Savannah, GA 31402-2206, joe.salamone@gulfstream.com)

A method is presented that computes a discrete-time model of the atmospheric turbulence effects on sonic booms. First, signal processing techniques are applied to sonic booms to extract the shock profiles from N waves measured on the ground. Next, clean shock profiles are estimated from nonlinear acoustics theory using the measured sonic boom ground data as a reference. A discrete-

time model is computed using a Wiener filtering scheme with the "clean" shock profile as the input to the scheme and the measured shock data as the output. The resulting system of equations required to compute a model is ill conditioned. Tikhonov regularization is implemented with the regularization parameter chosen based on inspection of the eigenvalue spectrum of the autocorrelation matrix. Multiple models are obtained for "low" to "high" amounts of turbulence from ground measurements recorded under various atmospheric conditions. Predicted N waves and shaped sonic booms are digitally filtered by the turbulence models. The loudness levels of the sonic booms before and after application of the turbulence models are evaluated.

11:00

2aPA9. Extension of a sonic boom model to ellipsoidal Earth and 3-D atmosphere. Kenneth J. Plotkin (Wyle Labs., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202, kenneth.plotkin@wylelabs.com) and Edward A. Haering, Jr. (NASA Dryden Flight Res. Ctr., Edwards, CA 93523-0273)

Sonic boom ray tracing models generally use flat-earth geometry, with 3-D ray propagation through a horizontally stratified atmosphere. There has been recent interest in long-range sonic boom propagation, both for analysis of sonic boom carpet edges from high-altitude flight and for analysis of over-the-top booms. The sonic boom program PCBoom4 has been extended to incorporate ellipsoidal earth and full 3-D ray tracing through a three-dimensional atmosphere. Ellipsoidal geometry of the Earth is handled by the method of Sofair [J. Guidance Control, **23**(2)], which defines a local vertical at given coordinates. A tangent reference plane is defined, and ray tracing is performed using the 3-D formulation of Schulten [NLR TP 97-374], in Cartesian coordinates. Once rays are traced, ray tube areas, signature aging, and focal zones are computed in the traditional manner. When over the top booms are analyzed, a starting signature for propagation after the caustic passage is generated from the recent numeric calculations by Kandil [AIAA Paper No. 2005-0013]. The new program, denoted PCBoom5, is expected to be available in 2007.

11:20

2aPA10. Supersonic aircraft: Shake, rattle and role. Sandy Liu (Federal Aviation Admin., Office of Environment and Energy, (AEE-100) Noise Div., Washington, DC)

There has been a renewed interest worldwide to focus on identifying technological advances that could satisfy the production of a new commercial supersonic aircraft. Although advances have been realized since the Concorde was developed, there is widespread agreement that many challenges remain. At a lesser risk, a supersonic business jet (SBJ) is seen as a technological stepping-stone in the overall tactic of advancing the speed of commercial aircraft. This presentation outlines some of the technological and environmental regulatory challenges facing such an aircraft. This includes a review of the established technical noise working group and activities that address supersonic aircraft environmental matters: sonic boom, aircraft community noise, and aircraft emissions. Federal Aviation Administration (FAA) maintains that a successful supersonic aircraft will depend on its stakeholders' ability to deal with these key issues. FAA finds it beneficial to work within the international arena for addressing these issues to ensure that all stakeholders' concerns are considered. Ultimately, successfully addressing concerns regarding sonic boom is a necessary initial condition for regulatory authorities to reconsider existing restrictions on supersonic operations, *the* key constraint to the business case in the revival to develop a civil supersonic aircraft for the future.

11:40

2aPA11. Flight demonstration project for low-sonic boom design technology. Akira Murakami (6-13-1 Osawa, Mitaka, Tokyo, 181-0015 Japan)

Japan Aerospace Exploration Agency (JAXA) has promoted a supersonic technology research program including flight demonstration since 1997. In this program, JAXA is now planning the new flight demonstration project "Silent Supersonic Technology Demonstrator Project (S3TD)." One of the targets of this project is to demonstrate the low sonic boom aircraft concept to reduce sonic boom intensity by 50% as compared with conventional configuration and to modify the sonic boom signature for not only front shock boom but also the rear shock one. JAXA will develop a new flight demonstrator that is an unmanned vehicle of approximately 3 tons in weight and 13 m in length, and then JAXA will make flight trials by 2012. Now JAXA is designing the base-line configuration, using the CAD-based Automatic Panel Analysis System that we have developed. The shock overpressure of the S3TD base-line configuration is estimated to be about 0.23 psf for front shock and 0.21 psf for rear shock, corresponding to about 50% intensity as compared with those of N-wave signature estimated by the first-cut method. The robustness of the low boom signature is also considered. The rear shock boom signature is so sensitive to flight condition that some small device is installed around the aft deck.

Session 2aPP

Psychological and Physiological Acoustics: Spatial Hearing and Temporal Processing (Poster Session)

Eric W. Healy, Cochair

Univ. of South Carolina, Dept. of Communication Sciences and Disorders, Columbia, SC 29208

Masashi Unoki, Cochair

JAIST, School of Information and Science, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

2aPP1. Motion perception of the target sound event under the discriminated two sound events. Tsuruyo Nishida (Gifu Natl. College of Tech., Motosu, Gifu 501-0495, Japan), Kazuhiko Kakehi, and Takamasa Kyutoku (Chukyo Univ., Toyota, Aichi 470-0393, Japan)

The correct percentage of motion direction for one sound event was about 90%. In the case of two sound events, the correct percentage was lower than for one sound event. The sound event was presented through the headphone in these experiments. These two sound events were physically separated both in frequency and in onset time for the two sound events. It might be possible to perceive two sound events separately. The perception of onset time difference and the discrimination of different frequency sound events were investigated to clarify whether the subjects perceived two sound events distinctly. When IEOI (interevent interval) was more than 40 ms, the subjects perceived the onset time difference and the order of two presented sound events. Results clarified that the subjects distinctly perceive two sound events separately with long IEOI. However, the correct percentage for the direction of motion perception of the target in two sound events was significantly lower than for one sound event. The target was strongly affected by the nontarget. Even when two sound events are perceived separately in terms of frequency and temporal order, it is difficult to perceive the motion direction of the target.

2aPP2. Perceptual movement of sounds moved up or down through multiway loudspeakers set up perpendicularly. Yu Agastuma and Eiichi Miyasaka (Musashi Inst. of Technol., 3-3-1 Ushikubo-Nishi, Tsuzuki-ku, Yokohama, Japan)

Perceptual movements of sounds fed through eight loudspeakers perpendicularly set up were investigated. The loudspeakers, which were mutually separated by 30 cm, were set up immediately in front of 15 observers positioned 4.5 m distant from the loudspeakers. They were required to answer the start point of the loudspeaker through which they perceived that the sound was output at the start, and the end point of that loudspeaker. The results show that the ascension or descension of the sound images was perceived when $\frac{1}{3}$ octave bands of noises with center frequencies of 1, 4, and 8 kHz were linearly moved up or down in the range from 60 to 150 cm. The results also show that the perceived moving distance depends on the moving speed, becoming longer than the physical distance at speeds of 10 and 20 cm/s, and converging to the physical distance concomitant with higher moving speed, irrespective of the center frequencies. [Work supported by SCAT.]

2aPP3. Combined manipulations of the perceived location and spatial extent of the speech-target image predominantly affect speech-on-speech masking. Ying Huang, Xihong Wu, Qiang Huang, and Liang Li (Speech and Hearing Res. Ctr., Peking Univ., Beijing, China, 100871, liangli@pku.edu.cn)

Speech maskers contain both informational-masking and energetic-masking components. To fully understand speech masking, it is critical to separate these two types of masking components. This study investigated the effect of the inter-target-source delay (ITSD) on intelligibility of speech when both the speech target and masker were presented by each of the two spatially separated loudspeakers (located at -45 and $+45$ degrees, respectively). The masker was either two-voice speech (different contents between the two loudspeakers) or steady-state speech-spectrum-noise (uncorrelated between the two loudspeakers). The results show that as the ITSD was decreased from 64 to 0 ms, the target image progressively became funneled into the region around the midline, and the intelligibility of the target was monotonically improved by over 40% when the masker was speech, but by only about 10% when the masker was noise. Under the quiet condition, however, the intelligibility was not affected by the change of ITSD. The results suggest that combined manipulations of perceived location and spatial extent of the speech-target image by changing the ITSD predominantly affect informational masking of speech. [Work supported by China NSF and Canadian IHR.]

2aPP4. Free-field sound localization with nonlinear hearing protection devices. Kimberly B. Lukas and William A. Ahroon (U.S. Army Aeromedical Res. Lab., Fort Rucker, AL 3636,)

Hearing enhancement and protection devices are being developed for both civilian and military markets that permit the wearer to maintain an awareness of the ambient sound environment and permit face-to-face communications while still protecting the user from high-level impulse noises or blast waves from, for example, handguns, rifles, and shotguns. The impulse noise protection for these devices can range from less than 9 dB peak pressure reduction to reductions of more than 40 dB peak. There has been little research, however, on the ability to use binaural cues to localize sound with such devices. An experiment to evaluate the ability of normal-hearing listeners to localize broadband noise bursts while wearing various nonlinear hearing enhancement and protection systems was conducted in an anechoic environment. Subjects were asked to localize a 60-dB SPL, 200-ms broadband noise burst from a speaker in 1 of 12 different azimuth locations (0 to 330 deg in 30-deg steps) and five different elevations (0,

± 15, ± 30 deg). The azimuth and elevation errors were measured for each device using circular statistical techniques. Front/back reversals were corrected. The results show a 15- to 20-deg degradation in the wearer's ability to localize sound when using these hearing protectors.

2aPP5. Selection of attractor sounds for an audio-based navigation system for the visually impaired. Brad S. Salisbury, Koorosh Naghshineh (Western Michigan Univ., Kalamazoo, MI 49008-5343), and William Wiener (Marquette Univ., Milwaukee, WI 53233)

Research into human hearing has been very laboratory oriented in the sense that the test environments do not replicate most real world situations [J. C. Middlebrooks and D. M. Green, Annu. Rev. Psychol. **42**, 135–159 (1991)]. While very useful information is gained from these types of tests, it is difficult to see how real world situations affect sound source localization, recognition, and navigation (walking/way finding) performance. Such information is especially important to people who are visually impaired and dependent on prior knowledge of the environment or audio cues for travel. The research reported here was conducted during the development of an audio-based navigation system. The question that arose was, what constitutes good attractor sounds versus bad sounds? A series of physical tests was developed to identify sounds that performed best from within a group of real-world attractor sounds. Testing was conducted with the aid of participants who were blind or visually impaired. Results of the experiments were consistent with those of Landau *et al.* [Asst. Technol. **17**, 133–143 (2005)].

2aPP6. A study of the equivalent perception between a visual stimulus with a feeling of depth and its associated sound. Hiroshi Hasegawa, Akimasa Suzuki, Ichiro Yuyama (Faculty of Eng., Utsunomiya Univ., 7-1-2 Yoto, Utsunomiya-shi, Tochigi-ken 321-8585, Japan), Hirotaka Ono, Masao Kasuga, and Miyoshi Ayama (Utsunomiya Univ., Utsunomiya-shi, Tochigi-ken 321-8585, Japan)

This study investigated the equivalent perception between a visual stimulus and its associated sound. Auditory-visual stimulus presentation experiments were performed using an audio-video clip of a man beating a drum on a road. The visual stimulus, with a perspective view of the road, imparted a depth sensation. Visual stimuli were projected onto a screen of 30.8 (W) (B 16.1 (H) deg. In all, 315 auditory-visual stimuli were produced with combinations of distance between the visual target (the drum) and the observer from (B00 m, the time delay between auditory and visual stimuli from −8F to 8F (1 F=$\frac{1}{30}$ s), and the sound pressure level of the sound stimulus from (B12 dB. The stimuli were presented to six subjects. In each presentation pattern, the sound pressure level to provide a perceptual strength equivalent to that of the visual target was determined: the equivalent sound pressure level (ESPL). Results show that the ESPL increased as the time delay between the auditory and visual stimuli increased, i.e., the ESPL tended to shift in the direction that the event was closer to the observer. The time delay shifted in the direction that the event was distant from the observer.

2aPP7. Localization of text-to-speech voices in a virtual speech display. Gordon Rubin (AT&T Labs, Florham Park, NJ 07932, and Rensselaer Polytechnic Inst., Troy, NY 12180)

An auditory display using high-quality text-to-speech (TTS) voices is presented to a group of individuals. Subjects are asked for their impression of the source direction of a synthesized talker in a nonreverberant environment. Location is coded using a set of nonindividualized head-related transfer functions (HRTFs). Many years of research in binaural technology provide comparative data for the localization ambiguities inherent in such presentation (e.g., front-back confusion). These ambiguities are characterized for stimuli generated by a mature concatenative speech synthesis system. An initial attempt is made to quantify the perceptual significance realized by a binaural-model-motivated speech synthesizer.

2aPP8. Using localization cues to track one of two concurrent sound sequences. Marie Rivenez, Anne Guillaume, Lionel Pellieux, and Christelle Delor (Departement Sci. Cognit. 5 IMASSA, BP 73, 91223 Bretigny sur Orge Cedex, France)

This study examined how localization, via binaural cues, can help the tracking of a sound sequence, interleaved with another sound sequence, across time. The ability to track a sequence was investigated using a temporal irregularity detection task. The temporal irregularity was inserted into a regular isotone sequence (target sequence) interleaved with an irregular sequence of same frequency (masker sequence). Performances were at the chance level when there was no interaural difference between the sequences. In experiment 1, large improvements of performances were found using several interaural level (ILD) or time (ITD) differences between the two sequences. Manipulating the masker sequence ILD and maintaining the signal-to-noise ratio constant in the best ear (experiment 2) or using a monaural presentation of the sequences (experiment 3), data showed that the effect of ILD could not be explained by a monaural processing (best or bad ear listening). Finally, experiment 4 revealed that the effect of binaural cues on tracking a sound sequence was mainly produced by the perceived localisation differences induced by the interaural differences rather than by the physical differences. These results suggested that binaural cues may influence sound sequence tracking via a spatial attention mechanism.

2aPP9. Comparison of pre- and post-cueing in a multiple-source sound localization task. Brian D. Simpson, Douglas S. Brungart (Air Force Res. Lab., 2610 Seventh St., Bldg. 441, Wright-Patterson AFB, OH 45433, brian.simpson@wpafb.af.mil), Robert H. Gilkey (Wright State Univ., Dayton, OH 45435), Nandini Iyer (Air Force Res. Lab., Wright-Patterson AFB, OH 45433), and James T. Hamil (General Dynam., AIS, Dayton, OH 45433)

Although most sound localization research has examined the ability of listeners to determine the location of a single sound presented in a quiet (typically anechoic) environment, most real-world listening situations are more complex, with multiple simultaneous sounds. Here, an initial experiment, designed to examine localization in multisource environments, is described. Listeners judged the location of a target signal (speech or environmental sound, presented normally or time-reversed) masked by up to four simultaneous sounds. In each block of trials, the observation interval was either preceded by, or followed by, a cueing interval, during which the stimulus to be localized was identified. It was expected that these two approaches would lead to different answers, as the associated tasks presumably address different listening strategies (i.e., analytic listening versus monitoring). The results indicate that, in all conditions, localization errors increase as the number of simultaneous sources increases. Moreover, performance degrades more rapidly in the post-cue condition than in the pre-cue condition. Surprisingly, this difference occurs for as few as two simultaneous sources, suggesting that there is a substantial cost when listeners are asked to remember what sounds were present and where those sounds were located in a complex auditory environment. [Work supported by AFOSR.]

2aPP10. Rate limitation and onset dominance in the discrimination of high-frequency click trains with dynamic envelope ITDs. G. Christopher Stecker (Dept. of Speech & Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105, cstecker@u.washington.edu)

Several studies have shown that binaural processing at high carrier frequency is limited to relatively slow modulations, suggesting profound consequences for binaural processing of high-rate electrical stimulation by cochlear implants (CIs). In this study, we measured normal-hearing-envelope ITD thresholds for Gaussian impulses (clicks) with 4000 Hz carrier frequency and 2 ms total duration (BW approximately 900 Hz) and for 16-click trains with rates of 100, 200, or 500 pps. For click trains, right-leading ITDs were static (condition RR), increased linearly from

diotic (condition 0R), or decreased linearly toward diotic (condition R0). Thresholds were calculated from the largest ITD in each stimulus (i.e., the onset of R0 or offset of 0R trains). Consistent with previous results for normal hearing and bilaterally implanted subjects, thresholds were strongly dependent on ICI. Low-rate (100 pps) thresholds improved for 16-click trains relative to single clicks in all conditions, while high-rate (500 pps) thresholds were similar to single clicks for RR and R0 stimuli. In condition 0R, 500 pps ITD thresholds were elevated and comparable to monaural rate discrimination. The results suggest an inability to lateralize high-rate stimuli in the absence of onset ITDs and have important consequences for binaural information processing with cochlear implants.

2aPP11. Evidence for active suppression of masker locations in spatially separated multitalker conditions. Kachina Allen, Simon Carlile, David Alais (Dept. of Physiol., Univ. of Sydney, F13 Univ. of Sydney 2006, NSW Australia, kachina_allen@yahoo.com), and Karen Froud (Columbia Univ., New York, NY 10027)

Three experiments tested active suppression of masker locations. Using a nonword corpus, spatial release from masking (SRfM) of 3.1 dB was calculated using masked thresholds (67%) of colocated and separated (20 deg) masker and target. Mirrored conditions were completed, with results pooled between hemispheres. In experiment 1, 80% of trials were masked 20 deg left of the central target. Random (10%) trials played targets at the expected masker location, and (10%) at the symmetrically opposite side, with maskers 20 deg laterally displaced. Thresholds revealed 2.6-dB reduction at the masker location compared to symmetrically opposite (null) locations. Experiment 2 used a central target with masker played 20 deg to the left in 80% of trials. Random (20%) trials (unexpected) played the masker at the symmetrically opposite side. Target thresholds were reduced by 2.1 dB when masker location was unexpected. Both experiments indicate active suppression of expected masker locations, suggesting masker suppression may contribute to SRfM. Experiment 3 played 80% of trials with maskers 20 deg left of the central target. Random (10%) trials played the target 40 deg to the left of center, and (10%) at the symmetrically opposite side, with the masker 20 deg laterally displaced. Reduced target thresholds (2.0 dB) indicated that masker suppression also extends laterally, beyond the expected masker location.

2aPP12. Lateralization of amplitude-modulated bone-conducted ultrasounds: Discrimination of interaural time and intensity differences. Takuya Hotehama and Seiji Nakagawa (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan, t-hotehama@aist.go.jp)

Ultrasonic vibration generates a sensation of sound via bone conduction; the phenomenon is called bone-conducted ultrasonic (BCU) hearing. Complex sounds also can be perceived by an amplitude-modulated BCU (AM-BCU). In this study, to investigate whether listeners can use an interaural time difference (ITD) in the amplitude envelope and an intensity difference (IID) as cues for lateralization, difference limens for an ITD (ITDL) and an IID (IIDL) for AM-BCU were evaluated. Thirty-kHz sine waves modulated by low-passed noises with cutoff frequencies of 250, 500, 1000, 2000, and 4000 Hz were used as the AM-BCU stimuli. ITDLs and IIDLs for air-conducted (AC) low-passed noises which correspond to modulator of the AM-BCU were also evaluated. Both ITDLs and IIDLs for AM-BCU were observed; AM-BCU was larger in ITDL and smaller in IIDL than AC. The results suggest that it is possible to lateralize AM-BCU using ITD in the envelope and/or IID; however, IID is a more effective cue than ITD.

2aPP13. Modulation masking in monaural and binaural conditions. Monika Kordus (Inst. of Acoust., Adam Mickiewicz Univ., Umultowska 85, 61-614 Poznan, Poland)

The shape of hypothetical modulation filters was measured in monaural and binaural signal presentation. A sinusoidal 4-kHz carrier was amplitude modulated by low-low-noise (LNN) [A. Kohlrausch et al., Acta Acust. **83**, 659–669 (1997)] and a probe tone of a frequency varying from 2 to 256 Hz. Using a three-interval forced-choice (3IFC) procedure the threshold of modulation by the probe tone was determined for the LNN of 16-Hz center frequency and 8-Hz bandwidth, and 64-Hz center frequency and 8-, 16-, or 32-Hz bandwidth. In the monaural condition, signals were presented to the right ear. In the binaural condition, modulation by the LNN was presented to the right ear, and modulation by the probe tone to the left ear. Results showed that masking in the modulation rate domain is characterized by a broad tuning with the highest values of about 17 dB for the monaural condition, and 4 to 9 dB in the binaural condition occurring for probe-tone frequency equal the center frequency of the LNN. The large effect in binaural condition confirms the idea that modulation detection by hypothetical modulation filter bank occurs on a higher level in the auditory system. [Work supported by the Polish Ministry of Education and Science, Grant No. 4T07B03230.]

2aPP14. Frequency dependency of illusory continuity in binaural hearing with respect to interaural time difference. Mamoru Iwaki and Hiroki Kitajima (Grad. School of Sci. and Technol., Niigata Univ., 8050 Ikarashi-2, Niigata, 950-2181, Japan, iwaki@bc.niigata-u.ac.jp)

Illusory continuity is a psychoacoustical phenomenon: when the absence portion of a tone is padded with sufficiently powerful white noise, an intermittent pure tone is perceived as though it were continuous. Many studies have examined this psychoacoustical illusion in monaural hearing, which supports that illusory continuity is observed when no clear evidence of discontinuity exists and when the power of the white noise is sufficiently large to mask the pure tone. In binaural hearing, the authors reported that illusory continuity was observed and that it was related with interaural time difference (ITD) for a single target pure tone of 500 Hz [M. Iwaki and N. Nakamura, INTERSPEECH/EUROSPEECH-2003 (2003) pp. 2065–2068]. This study investigated illusory continuity in binaural hearing for wider frequency ranges of 250–2000 Hz. Results showed that in the lower frequency range, the threshold level of illusory continuity in binaural hearing depended on ITDs between the tone target and noise masker. The increase of the threshold level was minimal when the target and masker have the same ITD. In the higher frequency range, the illusory continuity was independent of differences of ITDs.

2aPP15. Localizing noise in rooms via steady-state interaural time differences. Brad Rakerd (Dept. of Communicative Sci. and Disord., Michigan State Univ., E. Lansing, MI 48824, rakerd@msu.edu), William M. Hartmann, and Eric Pepin (Michigan State Univ., E. Lansing, MI 48824)

Noise lateralization experiments with headphones show that the steady-state interaural time difference (ITD) is, or is not, a valuable cue for sound lateralization depending on the height of the peak of the interaural cross-correlation function (cc). To discover whether these headphone results could be extended to the localization of sound in rooms, binaural recordings were made of one-third-octave noises in diverse environments: a dry classroom, a laboratory, and a reverberation room. Sound sources and a KEMAR recording manikin were moved about in each room to vary the cc from 0.2 to 0.8. The recordings, with ITDs digitally controlled, were presented to five listeners by headphones to measure listener sensitivity to ITD variations. It was found that sensitivity was almost entirely a function of cc with little weight attached to the actual room environment. ITD thresholds decreased monotonically with increasing cc. In agreement with earlier headphone studies using computed noises instead of room record-

ings, typical thresholds were higher by a factor of about 6 for a 2800-Hz band where envelope ITD is used compared to a 700-Hz band where waveform ITD is used. [Work supported by the NIDCD.]

2aPP16. Investigating the buildup of precedence effect using reflection masking. Jess Hartcher-O'Brien, M. Buchholz (Macarthur Auditory Res. Ctr. Sydney (MARCS), Bldg. 5, Bankstown Campus, UWS, Locked Bag, 1797, Penrith South DC), and M. Buchholz (Tech. Univ. of Denmark, DK-2800 Kgs. Lyngby, Denmark)

The auditory processing level involved in the build-up of precedence [Freyman *et al.*, J. Acoust. Soc. Am. **90**, 874–884 (1991)] has been investigated here by employing reflection masked threshold (RMT) techniques. Given that RMT techniques are generally assumed to address lower levels of the auditory signal processing, such an approach represents a bottom-up approach to the buildup of precedence. Three conditioner configurations measuring a possible buildup of reflection suppression were compared to the baseline RMT for four reflection delays ranging from 2.5–15 ms. No buildup of reflection suppression was observed for any of the conditioner configurations. Buildup of template (decrease in RMT for two of the conditioners), on the other hand, was found to be delay dependent. For five of six listeners, with reflection delay=2.5 and 15 ms, RMT decreased relative to the baseline. For 5- and 10-ms delay, no change in threshold was observed. It is concluded that the low-level auditory processing involved in RMT is not sufficient to realize a buildup of reflection suppression. This confirms suggestions that higher level processing is involved in PE buildup. The observed enhancement of reflection detection (RMT) may contribute to active suppression at higher processing levels.

2aPP17. Temporal processes involved in simultaneous reflection masking. Joerg M. Buchholz (CAHR, Acoust. Technol., Tech. Univ. of Denmark, Oersted Plads, Bldg. 352, DK-2800 Kgs. Lyngby, Denmark)

Reflection masking refers to the specific masking condition where a test reflection is masked by the direct sound. Employing reflection masking techniques, Buchholz [J. Acoust. Soc. Am. **117**, 2484 (2005)] provided evidence that the binaural system suppresses the test reflection for very short reflection delays and enhances the test reflection for large delays. Employing a 200-ms-long broadband noise burst as input signal, the critical delay separating these two binaural phenomena was found to be 7–10 ms. It was suggested that the critical delay refers to a temporal window that is employed by the auditory system to analyze the considered stimulus. In the present study, the dependency of this critical delay on the duration of the input signal (i.e., broadband noise) has been investigated. The psychoacoustical results revealed that the critical delay decreases with decreasing stimulus duration, resulting in a critical delay of about 2–3 ms for 20-ms-long stimuli. Hence, for very short stimuli the temporal window or critical delay exhibits values similar to the auditory temporal resolution as, for instance, observed in gap-detection tasks. It is suggested that the larger critical delay observed for longer stimuli is due to the auditory system focusing on lower signal frequencies.

2aPP18. Psychophysical estimates of cochlear tuning using simultaneous and nonsimultaneous masking. Masashi Unoki and Ryota Miyauchi (School of Information and Sci., JAIST, 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan, unoki@jaist.ac.jp)

This study attempts to estimate the tuning of an auditory filter, through both simultaneous and forward masking experiments with a notched-noise masker or a sinusoidal masker, to investigate how the effects of suppression affect estimates of frequency selectivity. In these experiments, masked thresholds of the short probe signal (5-ms ramps, no steady state) centered in the masker or delayed from the masker were measured for normal-hearing subjects as a function of the signal frequency (0.5, 1.0, 2.0, and 4.0 kHz), signal level (10, 20, and 30 dB SL), signal delay, and notch-width conditions in the notched-noise masker, or the masker fre-

quency in the sinusoidal masker. ERBs were calculated from derived auditory filters to compare the cochlear tuning. The estimated tunings became much sharper as the center frequency was increased. Moreover, the difference between the tunings of the filter at lower frequencies with either form of masking tended to be smaller than that at higher frequencies. The tuning may be affected by residual effects of suppression due to lower masker components below the signal frequency. [Work supported by special coordination funds for promoting science and technology of young researchers from the Japanese Ministry of Education.]

2aPP19. Practice effects in comodulation masking release. Cynthia M. Zettler and Marsha G. Clarkson (Dept. of Psych., Georgia State Univ., P.O. Box 5010, Atlanta, GA 30302, czettler@student.gsu.edu)

Substantial research has explored the acoustic factors influencing comodulation masking release (CMR), but the impact of practice has not been addressed. The current study evaluated the effect of experience on CMR thresholds. Ninety-eight adults listened to CMR stimuli through headphones once a week for 3 weeks. CMR was measured as the difference in threshold between a reference condition with one 20-Hz-wide amplitude-modulated noise band centered on a 1000-Hz, pure-tone signal, and a modulated masker condition having eight additional noise bands comodulated with the on-signal band. Thresholds were estimated in a single-interval, maximum-likelihood procedure. Hierarchical linear modeling tested the effect of practice on thresholds and CMR. Analyses indicated that thresholds improved significantly with practice ($p < 0.001$) for both masking conditions. On average, reference thresholds improved 5.41 dB per visit (SE=0.703), and modulated masker thresholds improved 3.15 dB per visit (SE=0.495). Total CMR decreased significantly ($p < 0.01$) by 1.98 dB per visit (SE=0.701), most likely because performance in the reference condition improved more than performance in the modulated masker condition. Results indicate that practice improves performance in CMR masking conditions and that total CMR decreases with experience.

2aPP20. Perceptual judgment in synchronization of two complex tones and its relation to the cochlear delays. Eriko Aiba (Grad. School of Music, Kyoto-City Univ. of Arts, 13-6 Oe-kutsukake-cho, Nishikyo-ku, Kyoto, 610-1197 Japan, m06901@kcua.ac.jp) and Minoru Tsuzaki (Kyoto-City Univ. of Arts, Japan)

Onset synchrony of components is widely assumed to be an important cue for perceptual unification as a single tone. However, even if the components begin simultaneously, their temporal relation might not be preserved at the cochlear level. The cochlear delay shifts the arrival of a lower component slightly but systematically behind a higher component. This experiment uses two complex tones as a stimulus to investigate whether the cochlear delay imposes a systematic bias in judging the synchrony of two complex tones. The results are compared with those of our previous experiment using two sinusoidal signals, which indicated that the accuracy of synchrony judgment was worse when the higher component preceded the lower one than in the opposite case. These results will be compared with those of a pioneering study, which indicated that the region of judgment on synchrony of complex tones was greater than 30 ms. [Work supported by the Grant-in-Aid for Scientific Research A, No.16200016, Japan Society for the Promotion of Science.]

2aPP21. Sequential amplitude-modulation interference across auditory channels and across ears. Magdalena Wojtczak, Melanie J. Gregan, and Neal F. Viemeister (Dept. of Psych., Univ. of Minnesota, 75 East River Rd., Minneapolis, MN 55455)

Forward masking has been demonstrated in the amplitude-modulation (AM) domain [M. Wojtczak and N. F. Viemeister, J. Acoust. Soc. Am. **118**, 3198 (2005)] for masker and signal imposed on the same carrier. This study explores an effect of AM on detection of AM imposed sequentially

on a different carrier. Data show that a 40−Hz AM (interferer) imposed on a 2−kHz carrier affects detection of a 40−Hz AM (signal) imposed on a 5.5−kHz carrier after the offset of the interferer AM. The effect is analogous (and similar in size for the shortest interferer−signal delays) to modulation detection interference for simultaneous interferer AM. In another condition, two independent noise samples were presented to opposite ears. In one ear, a 40−Hz masker AM was imposed on the carrier. In the opposite ear, a 40−Hz signal was imposed on the carrier after the offset of the masker modulation. A forward−masking effect similar to but smaller in size than that in monaural presentation was observed. The data from both experiments suggest that the forward−masking effect in the modulation domain occurs at central stages of processing, where information is combined across peripheral channels and where binaural interactions occur. [Work supported by NIH Grants DC00683 and DC006804.]

2aPP22. The influence of broadened auditory tuning on across-frequency integration of speech. Eric W. Healy and Kimberly A. Carson (Dept. of Commun. Sci. and Disord., Arnold School of Public Health, Univ. of South Carolina, Columbia, SC 29208)

Previous work has suggested that hearing-impaired (HI) listeners may have a deficit in their ability to integrate speech information across different frequencies. Performance of HI listeners was poorer than that of normal-hearing (NH) listeners on across-frequency tasks, and performance dropped more steeply when the signal was degraded through across-frequency asynchrony [E.W. Healy and S.P. Bacon, J. Speech Lang. Hear. Res. **45**, 1262–1275 (2002)]. In the current study, the role of broadened auditory tuning in this deficit was investigated. Broad tuning was simulated in NH listeners using noise-vocoded sentences and manipulating the slopes of the amplitude-modulated carriers. It was found that intelligibility of synchronous carriers was poorest in the broadest slope conditions, suggesting that poor integration of synchronous cues may be at least partially attributable to broad tuning. Performance reductions as a function of asynchrony were also found to be greatest in the broadest conditions, but only when analysis bands matched presentation bands and not when analysis bands were held constant. It is concluded that broadened auditory tuning may account for at least a portion of the across-frequency deficit observed in HI listeners. [Work supported by NIH.]

2aPP23. Effect of flanker band number and excitation pattern symmetry on envelope comparisons in masking release. Ifat Yasin (Dept. of Exp. Psych., South Parks Rd., Univ. of Oxford, Oxford OX1 3UD, UK), Deborah A. Fantini (Univ. of Essex, Colchester CO4 3SQ, UK), and Christopher J. Plack (Lancaster Univ., Lancaster LA1 4YF, UK)

Greater masking release (MR) is observed when narrow-band noise flanker bands (FBs) are lower, rather than higher, than the signal frequency. This may be due to more effective signal coding on the high-frequency side of the excitation pattern (EP) [B. C. J. Moore and U. Jorasz, J. Acoust. Soc. Am. **100**, 2373–2379 (1996)]. If so, masking of a high-frequency signal (asymmetric EP) should lead to a greater MR between FBs below and above the signal frequency than for a lower-frequency signal (symmetrical EP). Masked signal thresholds were compared for pure-tone signals (0.5, 1, or 2 kHz) masked by a narrow-band noise masker centered at the signal frequency [on-frequency band (OFB)] and two (three) FBs at 2-ERB intervals, below (above) the signal frequency. Masked signal thresholds were also measured in the absence of FBs. The OFBs and FBs were modulated at a modulation depth, m, of 0.7. FB envelopes were the same as each other and that of the OFB (comodulated, C), the same as each other but independent of OFB (incoherent, I), or not modulated (unmodulated, U). No clear relationship was found between EP symmetry and MR magnitude. MR is susceptible to the computation used, i.e., $[(I+U)−C]$ or $[I−C]$ as well as level cues.

2aPP24. Timing perception of tonal sequences connected with glides. Satomi Tanaka (Grad. School of Music, Kyoto City Univ. of Arts, 13-6 Kutsukake-chou, Oe, Nishikyo-ku, Kyoto-city, Kyoto 610-1197 Japan, m06902@kcua.ac.jp), Minoru Tsuzaki (Kyoto City Univ. of Arts), and Hiroaki Kato (ATR Cognit. Information Sci. Labs./NICT)

When we hear continuous sound signals, we extract the timing between each segment although each segmental boundary is acoustically unclear. In order to investigate on which points we perceive the temporal structures in sound signals, perceptual experiments were conducted to measure the sensitivity in detecting the temporal deviation by using sound signals containing frequency glides. The following question was raised: What would provide a more effective cue to detect a change in the temporal structures, the starting point or the ending point of glides? Five segments of steady parts, of which high and low were alternated, were connected with four frequency glides, while the amplitude was fixed. By randomizing the duration of glide parts and adjusting the duration of steady parts compemsatively, one can exclusively add a certain degree of fluctuation on the starting point or ending point of glides. The task was to detect a temporal deviation from an isochronous structure. The thresholds of detection were measured within three ranges of the center frequency. The detectability tended to be better under the starting point condition than under the ending point condition. The results were discussed in relation to a computational model of event detection. [Work supported by JSPS.]

2aPP25. Timing perception of sequences with no obvious onset: Isochrony detection in sequences of complex tones connected by formant glides. Minoru Tsuzaki (Kyoto City Univ. of Arts, NICT/ATR Spoken Lang. Commun. Labs., KCUA, 13-6 Kutsukake-cho, Oe, Nishikyo-ku, Kyoto 610-1197, Japan), Satomi Tanaka (Kyoto City Univ. of Arts, Japan), and Hiroaki Kato (ATR Cognit. Information Sci. Labs./ NICT)

To investigate the perceptual cues that detect arrivals of new auditory events, the detectability of deviation applied to an isochronous temporal structure was measured. The stimuli were sequences of complex tones whose formant frequencies alternate between two steady values. Five steady parts were connected by four formant glides. The isochronous structure was introduced exclusively in the intervals between the starting points of the glides or in the intervals between the ending points by adjusting the durations of the glides and the steady parts in a compensatory manner. Since the harmonic components stop and start one after the other during the glides, it is difficult to define the arrival of new events in an entirely physical view. The perceptual difference in detectability between the starting and ending conditions and its dependency on the frequency region could reveal the perceptual cues that are effective in extracting temporal relations of auditory events. The results will be compared with those obtained with sinusoidal stimuli whose frequencies were modulated in a similar way as the formant frequency in this paper, and they will be discussed by reference to a computational model of auditory event detection. [Work supported by JSPS No. 16200016.]

2aPP26. Human performance in detecting audio-visual asynchrony. Armin Kohlrausch, Steven van de Par (Philips Res. Labs. Eindhoven (WO02), HTC 36, NL-5656AE Eindhoven, The Netherlands, armin.kohlrausch@philips.com), Rob van Eijk (Techn. Univ. Eindhoven, NL-5600 MB Eindhoven, The Netherlands), and James F. Juola (Univ. of Kansas, Lawrence, KS 66045)

The perception of audio-visual delays by human observers is typically characterized by two quantities: the point of subjective equality (PSE) and the sensitivity to asynchrony. The PSE can be derived from a temporal order judgment (TOJ) procedure or from the outcomes of a synchronous-successive response paradigm. Sensitivity to asynchrony is derived from the steepness of the response curve in the TOJ paradigm. In this contribution we present data that show that PSE estimates derived from TOJ measurements are much more variable across observers than those based on synchronous-successive data. Two synchronous-successive methods were

used with different response categories: (1) two categories: asynchronous, synchronous, and (2) three categories: audio first, synchronous, video first. Both synchronous-successive methods yielded similar results. A slight influence of stimulus type on PSE estimates derived from synchronous-successive data was observed. To conclude, we analyzed discriminability values obtained with a two-alternative forced-choice procedure. Discriminability is better near the edge than in the middle of the synchronous response category. This suggests that categorical perception might play a role in audio-visual synchrony perception.

2aPP27. The effect of marker duration on time shrinking. Yoko Fujishima and Yoshitaka Nakajima (Faculty of Inst. of Design, Kyushu Univ. 4-9-1 Shiobaru, Minami-ku, Fukuoka, 815-8540, Japan, fujishim@gsd.design.kyushu-u.ac.jp)

Time shrinking is an illusion of time perception. This illusion takes place in stimulus patterns of three successive sound bursts (markers) of equal duration marking two empty time intervals, the preceding time interval (P) and the standard time interval (S). When S is longer than P, S is often underestimated. In this study, we examined whether the duration of the bursts (in this case, pure tones) could affect this illusion. Time shrinking always took place when $S-P$ was less than or equal to 60 ms, and the marker duration did not affect the amount of underestimation of S in this range. However, the illusion took place only for the shortest marker duration of 20 ms when $S-P=80$ ms. The results roughly agreed with previous data. [Work supported by JSPS and 21st Century COE, Kyushu University.]

2aPP28. The role of spectral factors in the occurrence of the gap transfer illusion. Tsuyoshi Kuroda, Yoshitaka Nakajima (Faculty of Design, Dept. of Acoust. Design, Kyushu Univ., 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan, kuroda@gsd.design.kyushu-u.ac.jp), Shimpei Tsunashima, Tatsuro Yasutake, and Shiro Mitsuba (Kyushu Inst. of Design, Minami-ku, Fukuoka 815-8540, Japan)

When a long ascending frequency glide with a gap at the temporal middle and a short continuous descending glide cross at their central position, observers report that the gap belongs to the short glide in spite of the physical position of the gap. This phenomenon is called the gap transfer illusion. We employed multiple-component crossing patterns and systematically varied the spectral slopes or formants of glides. In patterns that contained glides consisting of three harmonic components, the gap transfer illusion occurred when the crossing glides had equal spectral slopes, while the illusion was inhibited when the crossing glides had different spectral slopes. In patterns that contained 25 harmonic components representing formants of Japanese vowels, the gap transfer illusion took place only when the crossing glides had the same formants. Although Nakajima *et al.*

(2000) explained the gap transfer illusion in terms of the proximity of onsets and terminations, the present results show that it is also important to consider the role of spectral factors. [Work supported by JSPS and 21st Century COE, Kyushu University.]

2aPP29. An upper bound on the temporal resolution of human hearing. Milind N. Kunchur (Univ. of South Carolina, Dept. of Phys. & Astron., Columbia, SC 29208, kunchur@sc.edu)

This work obtained an upperbound of about 5 ms for the temporal resolution of hearing by studying human discriminability for bandwidth restriction by low-pass filtering. While low-pass filtering affects both the spectrum and temporal definition of the signal, the spectral changes in this experiment fall below the known just noticeable differences, so that the threshold seems to reflect the neural temporal-discrimination of the auditory system, rather than its ability to distinguish spectral cues based on tonotopic excitation patterns. The present result shows that human temporal resolution extends down to time scales much shorter than found in the past. Rough estimates of the expected resolution, based on convergencies in the cochlear nucleus and subsequent stages, yields values in a range that is consistent with the present experimental result.

2aPP30. Models of musical timbre using cortical spectro-temporal receptive fields and temporal codes. Mounya Elhilali (Inst. for Systems Res., Dept. of Elec. and Comput. Eng., A.V. Williams Bldg., Univ. of Maryland, College Park, MD 20742, mounya@isr.umd.edu), Daniel Pressnitzer (CNRS Université René Descartes Paris, 75005 Paris, France), and Shihab Shamma (Univ. of Maryland, College Park, MD 20742)

It has been customary to propose several "descriptive" definitions to quantify timbre as one of the basic attributes of sound. Multidimensional scaling techniques have proposed several acoustic dimensions to characterize the timbre of subsets of sounds. Yet, in the general case, timbre is an intricate percept difficult to relate to concise acoustic cues. Timbre probably reflects the dynamic interactions between spectral and temporal characteristics of sound that play a crucial role in identifying natural sources and convey a wealth of information about the environment. Recent physiological investigations revealed possible neural mechanisms to extract and represent the spectral and temporal information in complex sounds. At the level of the primary auditory cortex, spectro-temporal receptive fields exhibit a multiscale representation that captures many aspects of dynamic sound spectra. In the current study, we explore the potential correspondence between these cortical properties and the identification of musical instruments. We investigate the use of such representation as a quantitative descriptor of musical timbre, as validated by behavioral data from human subjects. Additionally, we examine biologically plausible coding strategies in this multidimensional cortical mapping involving spike-timing information.

Session 2aSC

Speech Communication: Prosody (Poster Session)

Victoria B. Anderson, Cochair

Univ. of Hawaii, Dept. of Linguistics, 1890 East West Rd., Honolulu, HI 96822

Keikichi Hirose, Cochair

Univ. of Tokyo, Dept. Information and Communication Engineering, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon.

2aSC1. The effects of speaking rate and tonal contexts on target undershoot of mandarin tone 3. Jia-Shiou Liao (Dept. of Linguist., Univ. of Texas, Calhoun Hall 501, 1 Univ. Station B5100, Austin, TX 78712-0198)

An investigation of the effects of speaking rate and tonal contexts on target undershoot of Mandarin tone 3 (low or low-falling-rising) was conducted. Data were from seven male and eight female subjects, all native speakers of Mandarin. The test phrase comprises three /ma/s in a row, which were put in a sentence to provide a natural context. The second /ma/ was always tone 3, preceded and followed by tone 1 (high-level), tone 2 (mid-rising), tone 3 or tone 4 (high-falling), resulting in 16 distinct combinations. Subjects read sentences at two relatively different rates: slow and fast. The fundamental frequency (F0) and duration of each /ma/ were measured at the onset and offset of its vowel. In addition, the duration of the second /ma/ and F0 at its midpoint were measured. Quantitative analysis of the rate of change in F0 suggests, by and large, that the F0 control systems are not always the same in slow and fast speech within the same tonal contexts. Furthermore, undershoot phenomena occur more often statistically across a syllable boundary than within a syllable boundary. Moreover, the F0 control system produces rates with less random fluctuation across the syllable boundary than within the syllable boundary.

2aSC2. Analysis on the effects of tonal coarticulation at word and nonword syllable boundaries of Mandarin based on the tone nucleus model. Qinghua Sun, Hirose Keikichi, and Minematsu Nobuaki (Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan, qinghua@gavo.t.u-tokyo.ac.jp)

As a typical tonal language, Mandarin has four lexical tone types, each of which has its own intrinsic pattern when uttered in isolation. However, when produced in context, the tonal contours present many variations, even if we specifically address the tone nucleus, which is defined as a stable part in a syllable. This study describes how and to what extent the word boundary affects tonal coarticulation in continuous speech. Although several previous works have investigated how the prosodic features of syllables are affected by their surrounding syllable tone types, most of them selected nonsense syllables uttered in a carrier sentence as study targets, which is far from natural situation. In this paper, after preparing a Mandarin speech corpus, in which all tone-type combinations of di-syllables are included with both intraword and interword cases, the degree of concatenation is compared for intraword and interword cases, viewing their tone component's F0 means and ranges of tone nuclei. All target di-syllables are uttered in a phrase. Results indicate that the degree in the intraword and interword cases differ greatly depends on the tone combination. The results were applied into our tone component prediction system, which is based on the tone nucleus model.

2aSC3. The command-response model for F0 contours and its application to phonetics and phonology of tone. Hiroya Fujisaki (Univ. of Tokyo, Tokyo, Japan)

The command-response model for the F0 contours, first proposed for nontone languages such an Japanese and English, has been extended by the author and his co-workers to cover F0 contour of tone languages and has been proved to be applicable to several dialects of Chinese including Mandarin, Cantonese, and Shanghainese, as well as other tone languages including Thai and Vietnamese. Namely, it allows one to approximate F0 contours of these languages/dialects with a very high accuracy using positive and negative tone commands. This paper first presents some results of analysis of F0 contours of these languages/dialects to demonstrate that the model can capture the acoustic-phonetic characteristics of their tones quite well. It then shows that coarse qualitative distinctions in the pattern of tone commands serve to distinguish tones within a language/dialect, as well as to distinguish tone systems of different languages/dialects. Thus the model proves to be a useful tool, not only in the studies of phonetic features of tones, but also in the studies of phonological structure of tone systems.

2aSC4. Tone sandhi effect on Chinese speech perception. Li Hsieh and Ya-Ju Yu (Dept. of Commun. Sci. and Disord., Wayne State Univ., 207 Rackham Hall, 60 Farnsworth, Detroit, MI 48202)

Mandarin Chinese has four tones with various sandhi processes. The third tone sandhi process is when a third tone (T3) becomes a second tone (T2) before another third tone (T3). Previous studies on third tone sandhi used connected speech in their experiments. This study examines the third tone sandhi effect on Chinese listeners in a nonconnected speech condition. Full-tone Chinese syllables were recorded in isolation. Ten Chinese listeners participated in tonal discrimination tasks which consisted of combinations of T2 and T3 (i.e., T22, T23, T32, T33). Three interstimulus intervals (ISIs) (i.e., 50, 250, 500 ms) and three types of pair conditions (minimal-, rhyme-, mixed-pairs) were introduced in the experiments. The findings of this study confirm the phonological hypothesis that the third tone sandhi effect has a great impact across all the pair conditions. When discriminating T23, Chinese listeners had longer response times and more errors than other tone pairs. The phonological phenomenon of third tone

3086   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3086**

sandhi persists regardless of connected or nonconnected speech conditions in Chinese spoken language. Differences in processing the tone sandhi pairs (T23) and the nontone sandhi pairs (T32) reveal the psychological reality of the third tone sandhi phonological rule.

2aSC5. A cross-language comparison of the use of stress in word segmentation. Michael D. Tyler (MARCS Auditory Labs., Univ. of Western Sydney, Bldg. 5, Bankstown Campus, Locked Bag 1797, Penrith South DC NSW 1797, Australia), Pierre Perruchet (Université de Bourgogne, Esplanade Erasme, 21000, Dijon, France), and Anne Cutler (Max Planck Inst. for Psycholinguist., 6500 AH Nijmegen, The Netherlands)

In spite of our illusions to the contrary, there are few acoustic cues to word boundaries in spoken language. While statistical probabilities between adjacent speech units provide language-general information for speech segmentation, this study shows that language-specific information may also play an important role. An $f0$ excursion, a correlate of stress, was provided on the first or last syllable of each word in an artificial language. As expected, French listeners succeeded with final-syllable stress but not initial-syllable stress. However, Dutch listeners appeared to benefit from both initial and final stress, rather than initial stress only. When the results were split according to the language background of the talker (French or Dutch), it became clear that the Dutch listeners only benefited from initial stress when the talker was Dutch, but not when the talker was French. Given that the majority of Dutch listeners spoke French, it is suggested that they readily adapted their word segmentation strategies when faced with French-sounding input (i.e., French talker or final stress). In support of this, Australian English listeners with no knowledge of French clearly benefited from initial stress only, consistent with English stress patterning. [Work supported by Conseil Regional de Bourgogne, NWO, CNRS, NIHDC00403.]

2aSC6. Acoustic and spectral correlates of lexical stress. Anthony O. Okobi (Speech and Hearing Bioscience and Technol., Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139)

Acoustic and spectral parameters that differentiate between primary and nonprimary stressed vowels were determined using two-syllable novel words with identical syllable compositions. The location of the high focal pitch accent within a declarative carrier phrase was varied using an innovative object naming task that allowed for a more natural and spontaneous manipulation of phrase-level accentuation. Results showed that when the high focal pitch accent was on the novel word, vowel differences in pitch prominence, intensity prominence, and vocal tract filter corrected amplitude of the first harmonic ($H1^*$) accurately distinguished primary from nonprimary stressed vowels. Vowel differences in first formant frequency, the spectral approximations of open quotient, $H1^*-H2^*$, and first formant bandwidth, $H1^*-A1^*$, also correlated with lexical stress under this condition. Correlates of lexical stress that were differentiating under all conditions tested were spectral tilt, $H1^*-A3^*$, noise at high frequencies, and duration. Furthermore, results indicated that these lexical stress vowel differences were augmented when the high focal pitch accent was on the novel word. This became apparent after a formula was devised to correct for the masking effect of phrase-level accentuation on the spectral tilt measurement. [Work partially supported by NIH T32-DC00038.]

2aSC7. The acoustic correlates of stress in compounds. Tuuli M. Adams (Linguist. Dept., New York Univ., 719 Broadway, 4th Fl., New York, NY 10003, tuuli.adams@nyu.edu)

The acoustic correlates of linguistic stress are traditionally considered to be intensity, pitch, and segment duration. The current study investigates the correlates of stress in English compounds such as ''blackboard'' by comparing their interaction with different intonational patterns. This study includes an experiment in which participants pronounced compounds in

controlled prosodic and intonational environments and various changes in intensity (dB), pitch (Hz), and duration were recorded for each vowel. Compounds in English are generally said to have a definitive stress pattern that distinguishes them from phrasal sequences; for example, ''greenhouse'' (for plants) would be stressed only on the first syllable, and its counterpart ''green house'' (house that is green) has a more even stress pattern. However, until now there has been no empirical study that successfully identifies the acoustic cues that distinguish compounds from phrasal structures. Results of this experiment show that intensity, pitch, and duration participate in a form of cue trading to signify stress in different contexts. This experiment provides new evidence that the acoustic correlates of stress interact with different intonational environments in a consistent way that distinguishes compounds from phrases.

2aSC8. Second occurrence focus and the acoustics of prominence. Jonathan Howell (Dept. of Linguist., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14850, jah238@cornell.edu)

Partee (1991) challenged the significance of the observation that certain adverbs (e.g., only) reliably associate with phonologically prominent words to truth-conditional effect, noting second occurrence (i.e., repeated or given) focus (SOF) appears to lack a phonological realization. Rooth (1996), Bartels (2004), Beaver *et al*. (2004), Jaeger (2004), and Fry and Ishihara (2005) argued that, while not intonationally prominent, an SOF word can be marked by increased duration and/or increased rms intensity. An acoustic study of verb-noun homophone pairs is reported. Three sophisticated speakers uttered five repetitions of the targets, embedded in discourses, in first occurrence (FOF), SOF, and unfocused (NF) contexts, both pre- and postnuclearly. Syntagmatic comparison shows a durational effect for focus and ordering in the FOF/NF case, yet only an effect of ordering in SOF/NF; similarly, an effect of spectral balance (H1-H3) was found for FOF/NF but none for SOF/NF. Results suggest, contrary to previous assumptions, that simple durational difference is not a straightforward cue of prominence. Sluijter and van Heuven (1996) and Campbell and Beckman (1997) have shown spectral balance calculated by various means is a cue to focus; its absence in these data raises doubts that prominence is reliably produced (let alone perceived) on SOF words.

2aSC9. Intonational distinctions between sentence types in North Kyungsang Korean. Hye-Sook Lee (Cornell Univ., 306 East State St., Apt. 305, Ithaca, NY 14850)

As a lexical pitch accent language, North Kyungsang Korean seems to display a very restricted distribution of intonation. This paper investigates intonational differences between statements and questions in North Kyungsang Korean. In particular, a variety of f0 targets are compared between the sentence types, including pitch accent f0 peak, utterance-final edge, pitch range, and f0 valleys. Besides the acoustic analyses, perceptual identification tests are performed to test whether apparently diverging acoustic cues are perceptually significant. Based on ANOVA tests, it turns out that NK Korean questions can be significantly different from statements with respect to the utterance-final edge f0 and the relative f0 and timing of the rightmost f0 valley. Specifically, questions have higher f0 values at the right edge and higher and earlier occurrences of the rightmost f0 valley. Further, the forced-choice identification tests, whose input materials are manipulated in the similar set of acoustic variables, show that only the utterance-final edge f0 is a strong perceptual cue to distinguish between questions and statements. In the meantime, the comparison of yes/no questions and wh-questions demonstrates that there is no common acoustic factor with which the two types of questions are significantly distinguishable.

2aSC10. The degree of word-initial low tone in Japanese: Syntactic boundaries and speech rate. Hisao Tokizaki (Sapporo Univ., ULS Sumikawa 701, Sumikawa 5-3-8-1, Sapporo, 005-0005 Japan)

This study investigates the nature of word-initial low tone in Japanese. It will be shown that the degree of low tones depends on the speech rate and the number of syntactic boundaries between the word and the immediately preceding word. It will be argued that the phenomena can be explained naturally by boundary mapping from syntax onto phonology (Tokizaki, 1999, 2006). Pierrehumbert and Beckman (1988) and Selkirk and Tateishi (1988) argue that minor phrases are demarcated by initial low tone in unaccented words. Selkirk *et al.* (2003) argue that degree of low tone depends on whether the word is located at the left edge of a maximal projection of the lexical head (XP). The analysis of spoken data will be presented. Radio news and recited sentences are analyzed by Praat. The pitches of the first and the second mora of a word are measured and compared with the pitch of the high tone in the preceding word. The pitch difference of the first mora and the second is calculated. The result shows that our analysis makes a better prediction for the degree of word-initial low tone than previous analyses. [This project is supported by Grant-in-Aid for Scientific Research, JSPS, 2006.]

2aSC11. Variability and mismatch: Speakers' cues and listeners' cues. Yuki Hirose (Dept of Lang. Information Sci., The Univ. of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan)

The present study questions the natural assumption that the types of cues speakers use are common to those that listeners utilize. The left-versus right-branching ambiguity (LB/RB) in Japanese complex NPs such as the blue CD-genitive case was the material of interest. In a production study, accompanying visual contexts were manipulated with respect to visual ambiguity: whether there was only one possible target (referent) object corresponding to either RB or LB in each scene, or objects corresponding to both analyses were present in each scene and therefore the ambiguity remains in the visual scene. Phonetic analyses of the collected utterances revealed that when the visual information eliminated the ambiguity, the difference between the LB- and RB-intended utterances was modest and appeared mainly in $F0$ cues; when the ambiguity remained in the visual context, the prosodic manipulation was made mainly on durational cues. Those utterances were then fed to a perception study. A multiple regression analysis indicated that the cues contributing to RB/LB distinctions in listeners mainly showed up on durational cues, suggesting that most types of cues manipulated by speakers who were not aware of the ambiguity did not contribute to listeners' comprehension.

2aSC12. A perception study of syntactic disambiguation using resynthesized speech. Mariko Sugahara (Doshisha Univ., Kyoto, Japan, msugahar@mail.doshisha.ac.jp) and Yuki Hirose (The Univ. of Tokyo, Tokyo, Japan)

It is already well known that when a preceding word (W1) is accented, the following word (W2) is likely to undergo a postaccent downstep. Such a downstep is especially evident in a left-branching structure (henceforth LB) as in [[W1 W2] W3], while it is mitigated in a right-branching structure (henceforth RB) as in [W1 [W2 W3]]. That is, ups and downs of the F0 peak of W2 disambiguate the RB and the LB structure when a postaccent downstep is available (Selkirk and Tateishi, 1991; Kubozono, 1993, Hirose, 2005). Two questions, however, emerge. One is whether or not the magnitude of W3's downstep would ever contribute to the branching disambiguation. The other is what would happen when all words are unaccented, i.e., when no postaccent downstep is available. We carried out a perception study using resynthesized speech as stimuli, to examine those questions. Our results show that neither the W3 peak values of a postaccent environment nor the W2 peak values of an unaccented environment had significant effect on the LB vs. RB disambiguation. Such results will be compared with Hirose's (2005) production and perception study on spontaneous speech.

2aSC13. The effect of segmental structure on the production of Japanese pitch-accent. Irina A. Shport (Linguist. Dept., Univ. of Oregon, 1290 Univ. of Oregon, Eugene, OR 97403, ishport@uoregon.edu)

Japanese pitch-accent is a fall in fundamental frequency ($F0$) after the accented mora. In current models of Japanese pitch-accent contrast, accented and unaccented pitch patterns are predicted to be similar across syllable structure type. This study provides evidence that the segmental make-up of words affects accentual patterns in Tokyo Japanese in terms of their $F0$ contours. Six native speakers produced accented and unaccented words that varied in segmental type of the word-medial mora ($N = 864$). It was found that the pitch-accent contrast was produced differently in words with a short vowel in the first syllable, CVCVCV and CVC:V words (V-group), on one hand, and words with a longer voiced portion of the first syllable, CV:CV and CVNCV words (V:/VN-group), on the other hand, $p > 0.01$. In the accented condition, the average $F0$ fall was 50 Hz for all, but the $F0$ movement within the first syllable was different between the two word groups. In the unaccented condition, word-level, as well as intrasyllabic differences, were observed in $F0$ contours. $F0$ was rising in the V-group, whereas it was flat in the V:/VN-group. These results indicated that pitch patterns varied by segmental structure, more so for unaccented words.

2aSC14. Perception of Japanese pitch accent: Examination of minimal accent pairs. Yukiko Sugiyama (Dept. of Linguist., Univ. at Buffalo, SUNY, 609 Baldy Hall, Buffalo, NY 14260, ys8@buffalo.edu)

Perception of Japanese pitch accent was examined using 20 bimoraic/disyllabic minimal accent pairs. A minimal accent pair refers to a pair of words that is identical except for their accent types. For example, /hana/ "flower" and /hana/ "nose" have the same phonemic sequence and the pitch pattern of low-high yet they differ in that "flower" has accent on the word-final mora while "nose" has no accent (unaccented). A preceding production study that utilized the same 20 pairs of words found that the maximum F0 in the word-final mora was reliably higher for the final-accented words than the unaccented words and the F0 rise from the initial mora to the final mora was reliably larger for the final accented words than the unaccented ones when they were produced sentence-medially followed by a particle. However, the two accent types did not show any reliable difference when they were produced by themselves in isolation. This perception study accesses if the properties found to be related to Japanese pitch accent in production are actually used by listeners when they identify accent types. The relation between the production and perception of Japanese pitch accent will be discussed.

2aSC15. The phonetic and phonological triggers for downstep in Japanese: Evidence from perception. Takahito Shinya (Dept. of Linguist., 226 South College, Univ. of Massachusetts, Amherst, MA 01003, tshinya@linguist.umass.edu)

In Japanese the F0 peak of a word is downstepped after a lexically accented word. Lexical accent can be characterized by its phonological specification in the lexicon or by its phonetic F0 pattern. A perceptual experiment was conducted in order to observe how these two properties interact with each other as downstep triggers. The experiment tested whether listeners compensate for downstep exclusively based on the phonological information of lexical accent or on the phonetic cues available. Participants were asked to judge the relative prominence between two F0 peaks. Results suggest that in order for two words to be perceptually equal in prominence, the second F0 peak must be lower than a given first F0 peak. This difference between the two F0 peaks was significantly greater when the first word was phonetically accented than when it was phonetically unaccented, which suggests that the downstepped F0 peak is perceptually compensated. A similar effect was observed for the phonologically accented first peak, but the effect size was much smaller than the effect observed for the phonetically accented first peak. These results suggest that the primary trigger of downstep in Japanese is the F0 pattern of accented words. Lexical specification plays, at best, a secondary role.

2aSC16. Pitch values in prominent and nonprominent sentence enders in Japanese. Yumiko Enyo (Dept. of Linguist., Univ. of Hawaii at Manoa, 1890 East-West Rd., Honolulu, HI 96822)

This study uses naturally occurring interview data to investigate properties of intonational prominence in Japanese sentence enders (SEs). Since Japanese intonation patterns generally observe downstep (Venditti, 2005), a sentence-ending AP (accentual phrase) is likely to have the lowest F0 in an utterance if it does not receive prominence. Data from one interviewee's responses, SEs in long IPs (more than 5 APs) and short IPs (one or two APs), are used for this analysis. Within the group of 28 short IPs, there were eight cases of the SE nandaroo "what is it, I wonder." Measurements of highest and lowest F0 within nandaroo indicate two distinctive F0 patterns, determined by whether the SE is prominent or nonprominent. (Prominence appears to be placed on an SE when its content is focused.) In prominent SEs, the intonation contour contains peaks that exceed three standard deviations of the speakers F0 average. However, prominent and nonprominent utterances are similar in their F0 minima. As a result, SEs with prominence typically show wider pitch range than those without prominence. Similar results are found for other SEs contained in both long and short IPs.

2aSC17. On the prosody control characteristics of nonverbal utterances and its application to communicative prosody generation. Ke Li, Yoko Greenberg, Nagisa Shibuya, Yoshinori Sagisaka (GITI Waseda Univ. 29-7 Bldg. 1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo 169-0051, Japan,rikoku@akane.waseda.jp), and Nick Campbell (NiCT/ATR)

In this paper, prosodic characteristics of nonverbal utterances were analyzed using an $F0$ generation model proposed by Fujisaki aiming at communicative speech generation. From the analysis, the different distributions of $F0$ generation parameters have been observed for prototypical four dynamic patterns (rise, gradual fall, fall, and rise&down). Since former works have shown that these differences can correspond to their impressions (such as confident-doubtful, allowable-unacceptable, and positive-negative) expressed by multi-dimensional vectors, we tried to make a computational model from impression vector to $F0$ generation parameters. By employing a statistical optimization technique, we have achieved the mapping from impression vectors to prosody generation parameters. Perceptual evaluation tests using neutral words have confirmed the effectiveness of the mapping to provide communicative speech prosody. [Work supported in part by Waseda Univ. RISE research project of "Analysis and modeling of human mechanism in speech and language processing" and Grant-in-Aid for Scientific Research B-2, No. 18300063 of JSPS.]

2aSC18. Analyses of fundamental frequency contours for Japanese utterances that express several degrees of four emotions—joy, sadness, anger, and fear. Hiromi Kawatsu and Sumio Ohno (Grad. School of Bionics, Comput. and Media Sci., Tokyo Univ. of Technol., 1404-1 Katakura, Hachioji, Tokyo, 192-0982, Japan, kawatsu@so.cs.teu.ac.jp)

To give synthetic speech richer expression, prosodic features of utterances with various kinds of emotions were analyzed. Utterances that express four basic emotions with several degrees were collected as speech material: joy, sadness, anger, and fear. The fundamental frequency contours are analyzed based on a model for the process of generation. Changes in controlling parameters of the model were examined with regard to degrees of emotion. The baseline frequency increases as degrees of emotion increase. Especially for sadness and anger, its tendency is remarkable. Regarding phrase commands, the rate of occurrence increases as the respective degrees of emotion increase. The rates are affected by the kind of branch boundary in the grammatical structure and the number of morae from the immediately preceding phrase command. The change of the amplitude of phrase commands depends on the kind of position of grammatical structure. For the accent commands, timings of their onsets and offsets

are almost constant for degrees of emotion. They depend on the accent types of prosodic words. The magnitude of the accent commands changes as degrees of emotion increase depending on the positions of prosodic words from the beginning of the utterance.

2aSC19. An acoustic functional grammar of English intonation. Hunter B. Hatfield (Univ. of Hawaii, 438 Launiu St., Apt. A, Honolulu, HI 96815, hunterh@hawaii.edu)

There has been a resurgence in recent years of linguistic models that directly embed exact physical and acoustic data right into the formalism. This work attempts to show that many aspects of speech perception and production can only be understood when the acoustic input is fully considered. In the case of speech perception, if there are abstractions, they are abstractions over acoustic input, not based on *a priori* phonological categories. The current paper extends this work into English intonation. The functional phonology of Paul Boersma, which incorporates exact frequency values into an optimality theoretic perception grammar, is extended to explain several basic facts about English intonation, including the derivation of categorical high and low tones from the acoustic signal, the relation of tones to hierarchically adjacent tones, and the affectual aspects of classically paralinguistic intonation. The primary advantage of this approach is that it allows a single model to handle both linguistic and paralinguistic intonation structure, a merger which is only possible due to the use of physical acoustic data in the model.

2aSC20. Does English have accent domains? Molly Shilman (Linguist. Dept., UCLA, Los Angeles CA 90095-1543, mshilman@humnet.ucla.edu)

In ToBI systems, there are languages in which a pitch accent is the head of a particular level of phrase (accentual phrase) (e.g., French, Japanese), and those in which pitch accents are largely independent of phrasing (e.g., English, German). However, some theories of syntax-prosody mapping, such as Truckenbrodt's (1999), assume that a pitch accent is always the head of some prosodic domain. Such theories thus predict that every language with pitch accents should have a prosodic unit that is the domain of pitch accent. Previous work suggests that prosodic boundaries tend to be marked by several types of phonetic cues, including final lengthening (Wightman *et al.*, 1992; Jun, 2005). This study uses a controlled production experiment to determine whether there are accent domains in English: speakers produce near-minimal triplets in which the syntactic relations between words are changed by context but the segmental string, location of word boundaries, and location of the single pitch accent within the target phrase all remain fixed. Productions are then compared for final lengthening at the location of target boundaries. Preliminary results suggest that there is an accent domain in English.

2aSC21. Testing the implicit prosody hypothesis: Data from English. Sun-Ah Jun (Dept. of Linguist., UCLA, Los Angeles, CA 90095-1543, jun@humnet.ucla.edu)

One of the recent theories in sentence processing hypothesized that the prosodic phrasing implicit to a certain syntactic structure determines how speakers interpret the structure. According to this implicit prosody hypothesis (IPH; Fodor, 1998, 2002), in a structure of relative clause (RC) with a complex head noun, speakers/languages tend to prefer low attachment (i.e., an RC is modified by the adjacent head noun) if the prosodic break adjacent to RC is smaller than that between the head nouns. However, speakers/languages tend to prefer high attachment (i.e., an RC is modified by the nonadjacent head noun) if the prosodic break adjacent to RC is greater than that between the head nouns. Jun and Koike (2003) and Jun and Kim (2004) have shown that Japanese and Korean speakers who prefer high attachment produced a bigger prosodic break adjacent to RC than between the head nouns, confirming the IPH. The current study examines the prosodic phrasing of English sentences to investigate whether speakers

of American English, known to prefer low attachment, would produce a larger prosodic break between the two head nouns than that between the RC and the adjacent head noun. Data from 30 speakers will be discussed. [Work supported by UCLA Faculty Grants.]

2aSC22. Phrasal length and complexity effects on interphrase pause duration. Jelena Krivokapic (Dept. of Linguist., Univ. of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693,krivokap@usc.edu)

This study examines the effect of prosodic structure and phrase length on pause duration and the speech planning processes that such pauses can reveal. Subjects read sentences varying in two ways. First, each sentence included a pause bounded either by long intonational phrases both before and after it, medium phrases before and after, or short phrases before and after. Second, the prosodic complexity of the phrase on each side of the pause, i.e., before it and after it, was varied as to whether it was a branching intonational phrase or a nonbranching intonational phrase. Fourteen speakers read sentences synchronously in dyads so as to minimize variability due to speech rate and individual differences [Cummins, ARLO (2002), Cummins, J. Phonetics: (2003); Zvonik and Cummins, ICSLP (2002)]. The results for the dyads analyzed to date indicate that preboundary branching phrases are followed by longer pauses than nonbranching preboundary phrases. Further, surrounding long phrases lead to longer pauses than surrounding medium phrases, which in turn give rise to longer pauses than surrounding short phrases. Interaction effects of phrasal length and complexity will also be discussed. [Work supported by NIH.]

2aSC23. Speech timing: Final lengthening and syllable reduction in cerebellar pathology. Maureen A. Casper and Jennifer Geibel (State Univ. of New York @ New Paltz; 75 South Manheim Blvd., New Paltz, NY 12561)

The production of prosodic prominence may be considered the hallmark of normal speech motor control. Two theoretical views predict different relations between duration and formant and fundamental frequencies. The dynamic view of coordination predicts a disassociation between time and movement amplitude, while the clock theory of cerebellar movement poses a monotonic relation between duration and frequency. Phrase-final lengthening and syllable reduction each were measured across a prosodic contrast design for the acoustic measures of syllable duration, $F0$, $F1$, and $F2$. The speech characteristics of normal and cerebellar ataxic speakers were compared. In support of gestural dynamics, and following the kinematic findings for normal speakers, the acoustic findings showed longer duration and reduced frequency for phrase-final lengthening, and shorter duration and reduced frequency for syllable reduction. Cerebellar speakers produced longer durations for both events and increased variability for formant and fundamental frequency. This dissimilarity between normal prosodic production and that of cerebellar speakers implies a dynamic relationship between time and articulatory movement. In contrast to normal speech production, cerebellar speech is markedly slow and fails in the rapid adaptation to reduce frequency in phrase-final and reduced syllables.

2aSC24. Voice $F0$ responses to actual versus consciously perceived auditory feedback pitch shifts. Theresa Burnett and Dana Madonia (Dept. of Speech & Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

The purpose of this study was to determine the effect of pitch shift discrimination ability on production of voice $F0$ responses to auditory feedback pitch shifts. Subjects repeatedly vocalized /u/ into a microphone while listening to their voice fed back over headphones. During each trial, auditory feedback pitch was experimentally shifted upward or downward (randomly determined) by 50 cents for 100 ms. Subjects subsequently pressed a button labeled up or down to indicate their perceived direction of

voice feedback pitch shift. Voice output was recorded, $F0$ extracted, smoothed, time aligned to the onset of the pitch shift, and averaged. Button response accuracy was found to be at chance expectancy among the majority of subjects. Nevertheless, all subjects responded vocally to the pitch shifts by changing voice $F0$ in the opposite direction, indicating that the direction of the shift was indeed processed by the auditory system. These results are consistent with the existence of preattentive mechanisms of audio-vocal control, and provide evidence for a mismatch between conscious and preattentive auditory processing of voice pitch feedback for the control of voice $F0$. [Work supported by NINDS.]

2aSC25. Correlations between voice $F0$ responses and manual reaction times to auditory feedback pitch shifts. Dana Madonia and Theresa Burnett (Dept. of Speech & Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

There appear to be two types of voice $F0$ responses to voice auditory feedback pitch manipulations: early, opposing responses, and later responses that vary across subject and task. The purpose of this study was to determine whether either type correlates with volitional motor abilities. Subjects vocalized /u/ into a microphone while listening to their voice fed back over headphones. Subjects also pressed a joystick button as rapidly as possible upon hearing any experimenter-generated auditory feedback pitch shift. Pitch shift stimuli were 50 cents upward or downward (randomized), and 100-ms duration. Voice was recorded, $F0$ extracted, smoothed, time aligned to the onset of the shift, averaged according to shift direction, and the onset of the first voice $F0$ response was measured. Average manual reaction time was calculated. Findings indicated no significant correlation between manual responses and early voice $F0$ response. However, later voice $F0$ responses and manual reaction times were positively correlated. These results are consistent with the hypothesis that the early voice $F0$ response is automatically generated, and later responses are under voluntary influence. Results also suggest that the variability of later voice $F0$ responses may be a function of individual differences in volitional motor response abilities. [Work supported by NINDS.]

2aSC26. Frequency and amplitude derivatives as syllable-level $F0$ features. Mark Liberman and Jiahong Yuan (Univ. of Pennsylvania, Philadelphia, PA 19104)

Most descriptions of tone and intonation focus on the alignment of $F0$ rises, falls, and inflection points relative to the phonetic unfolding of associated syllables. In this context, it is suggested that the relationship between proportional changes in $F0$ (treated modularly to eliminate the effects of period doubling and halving) and proportional changes in amplitude (within the frequency bands typically dominated by energy from voicing) provides an interesting characterization of $F0$ trajectories in relation to syllabic structure. Using data from conversational and broadcast speech in English and Chinese, parameters of this type (in combination with other acoustic measures) are used in identifying relevant features of tone and intonation.

2aSC27. Declination of subglottal pressure during speech and its relation to tone distribution. Helen M. Hanson, Janet Slifka, Stefanie Shattuck-Hufnagel (Speech Commun. Group, Res. Lab. of Electron., MIT, 77 Massachusetts Ave., Cambridge, MA 02139), and James B. Kobler (Massachusetts General Hospital, Boston, MA 02114)

The subglottal pressure contour P_s for speech is considered to have three phases: initiation (rapid rise), working (level or slightly declining), and termination (rapid fall). The current work focuses on characterization of the working phase in terms of the distribution of pitch accents and of phrase and boundary tones. In particular, the degree of P_s declination is studied. A measure of P_s declination has proven difficult to define [cf. Strik and Boves, J. Phonetics **23**, 203–220 (1995)]. Therefore, in pilot work, subjective ratings of degree of declination are made on a subset of

a corpus in which the tone distribution is controlled. Significant variation in the degree of declination is observed among speakers. For example, P_s for one speaker is relatively constant to slightly declining, while for another it is almost always sharply declining. For some speakers utterances with an early-occurring nuclear pitch accent (NPA) show greater degree of declination than utterances with a late-occurring NPA, as do utterances ending with high tones. Data for additional subjects and objective measures will be presented. The results have implications for models of speech production, and for applications such as computer speech synthesis and recognition. [Work supported by NIH Grants DC00075 and DC04331.]

2aSC28. Comparing methods for measuring pitch "elbows." Ryan Shosted,[a)] Alex Del Giudice, and Amalia Arvaniti (Dept. of Linguist., Univ. of California, San Diego, 9500 Gilman Dr., #0108, La Jolla, CA 92093)

Currently, there is no standard method for determining the location where an F0 curve begins its upward or downward trajectory (sometimes referred to as the pitch "elbow"). It is generally accepted that elbows are hard to measure in a consistent way: some studies have used least-squares fitting methods, others have relied on visual judgments. This study uses a variety of data from English and Greek that exhibit different degrees of difficulty for locating the elbow. Automated results are derived from various implementations of two algorithms. The first algorithm uses linear least-squares fitting, where the intersection of two lines that best approximate the pitch contour denotes the point of the elbow. The second algorithm analyzes point-by-point changes. The results are then compared with those of human labelers. We present results from a series of statistical tests that estimate the robustness of variations on these two algorithms with respect to the performance of the human labelers. Our aim is to provide a method for measuring elbows that is robust and easy to implement and that can be used with different types of pitch contours, in order to improve accuracy of measurements. [a)] Also at Univeristy of California, Berkeley.

2aSC29. Semi-automatic labeling of pitch accents in American English. Diana Stojanovic and Yohei Sakata (Dept. of Linguist., Univ. of Hawaii, 1890 East-West Rd., Honolulu, HI 96822)

This study provides a simple classification of several pitch accents that occur in American English, based on acoustic measurements of F0. Audio samples recorded by a female speaker and transcribed in the ToBI framework (Beckman, 1994) were used to develop five rules to recognize the pitch accents H*, L*, L+H*, and L*+H. Since no information about prosodic phrasing was given to the recognizer, some combinations with phrase accents were considered and contours then grouped into equivalent labels (e.g., %HL* equivalent to H*L−). Based on the rules developed, a semi−automatic process of labeling pitch accents was implemented in Praat. Half of the audio samples were used to train the labeler, the other half to test it. The accuracy level achieved was 69%. When effects of microprosody, creaky voice, or incorrect automatic pitch tracking were removed, the accuracy of classification into equivalent labels increased to 90%. Two types of applications are proposed: semi−automatic computer labeling of intonation for larger quantities of data, and a tool to help human transcribers learn American English ToBI using Praat. In addition, the simple classification is the first step towards looking into speaker variability in realization of different contours, and typology of intonational contours in different languages.

2aSC30. Parametric spectrogram modeling of single and concurrent speech with spline pitch contour. Jonathan Le Roux, Hirokazu Kameoka, Nobutaka Ono, and Shigeki Sagayama (Grad. School of Information, Phys. and Computing, The Univ. of Tokyo. 7-3-1, Hongo, Bunkyo-ku 113-8656 Tokyo, Japan)

This paper proposes a novel general speech analysis framework consisting in a precise and parametric description of speech directly from the power spectrum, resulting in a new efficient pitch estimation method for both single and multiple speaker situations. The speech spectrum is modeled as a sequence of spectral clusters governed by a common pitch contour expressed as a spline curve. These clusters are isolated by an unsupervised 2-D time-frequency categorization of the energy density, and their common pitch contour is estimated simultaneously. A smooth pitch contour can thus be extracted on the whole utterance, linking together its voiced parts, while this is difficult or impossible for frame-by-frame algorithms. The proposed algorithm is compared to several existing methods on a database of single speaker speech, with very good results that place it among the most accurate ones. Moreover, its very high accuracy is shown on the estimation of multi-pitch contours of co-channel concurrent speech.

2aSC31. Robust $F0$ estimation based on complex speech analysis. Tatsuhiko Kinjo and Keiichi Funaki (Univ. of the Ryukyus, Okinawa, Japan)

This paper proposes a novel robust fundamental frequency ($F0$) estimation algorithm based on complex-valued speech analysis for an analytic speech signal. Since the analytic signal provides spectra only over positive frequencies, spectra can be estimated accurately at low frequencies. Consequently, it is considered that $F0$ estimation using the residual signal extracted by complex-valued speech analysis can perform better for $F0$ estimation than that for the residual signal extracted by conventional real-valued LPC analysis. In this paper, the autocorrelation function (AUTOC) weighted by a reciprocal of the AMDF is also adopted for the $F0$ estimation criterion. The proposed $F0$ estimation algorithm is evaluated using three criteria, AUTOC, AMDF, and the weighted AUTOC, with complex-valued residual extracted by complex-valued speech analysis. We also compared the proposed method with those for three signals: the speech signal, analytic speech signal, and LPC residual. Speech signals used in the experiments were corrupted by addition of white Gaussian noise whose noise levels are 10, 5, 0, and −5 [dB]. The performance was evaluated using the gross pitch error (GPE) and fine pitch error (FPE). The results demonstrate that this algorithm based on complex speech analysis can perform better than conventional methods in an extremely noisy environment.

2aSC32. Cross-linguistic examination of listeners frequency thresholds for musical tones and vowels. Daniel Pape (ZAS Berlin, Jgerstrasse 10-11, 10117 Berlin, Germany)

There is clear evidence that musical education increases the discriminability of the frequency threshold (dl). If musical education increases sensitivity, the question arises if listeners of a language that uses F0 for prosody are more sensitive to musical tones and vowels' dl thresholds. A perception experiment was conducted examining the dl of musical tones and vowels for both professional musicians and nonmusicians. The experiment was carried out cross linguistically: The languages include German and English (Germanic languages) and Italian, Portuguese, and Spanish (Romance languages). Results indicate that dl is significantly smaller for musicians compared to nonmusicians for all examined languages. The native vowel dl is significantly smaller for the Germanic languages. This is attributed to the more frequent use of F0 for prosody, thus claiming linguistic reasons. Surprisingly the dl for the musical tone in all languages is smaller for the Germanic musicians compared to the Romance musicians, although musical education and the stimulus were identical for all musically educated musicians. These results indicate an influence of the native language on the pitch accuracy of the listeners. Thus we give evidence that interference of speech pitch and musical pitch processing occurs, which supports the "common processing theory."

2aSC33. Peak alignment in accented Tolowa syllables. Christopher Doty (Dept. of Linguist., Univ. of Oregon, 1290 Univ. of Oregon, Eugene, OR 97403-1290, suomichris@gmail.com)

Tolowa (Athabaskan, Pacific Coast Group) has only three remaining fluent speakers near Crescent City, CA. Tolowa has prominent pitch contours that have been described both as an interaction of stress and vowel length and as lexically specified pitch accent. The present study examined

the phonetic realization of these pitch contours. Fundamental frequency ($F0$) at its highest peak during the rhyme of accented syllables was measured, as was relative duration into the rhyme at which the peak occurred. Preliminary results indicated that a high pitch peak occurred approximately 20%–30% of the way into the rhyme of the syllable. In the case of long vowels in accented syllables, this led to a prominent peak which fell sharply during the second half of the rhyme. A similar peak and fall was seen in accented syllables closed by a voiced sonorant. In the case of accented syllables with short vowels closed by a voiceless consonant, the accent was realized as a rise in pitch leading to the final consonant. These results indicate that the phonetic target is a high peak 20%–30% into the rhyme; the target seems to not include the fall seen in rhymes with long vowels or those closed by voiced sonorants.

2aSC34. Production characteristics of accent in four dialects of Nahuatl. Susan G. Guion (Linguist. Dept., Univ. of Oregon, 1290 Univ. of Oregon, Eugene, OR 97403, guion@uoregon.edu), Jonathan D. Amith (Gettysburg College, Gettysburg, PA 17325), Christopher Doty, and Irina A. Shport (Univ. of Oregon, Eugene, OR 97403)

Most descriptions of Nahuatl (Uto-Aztecan, Mexico) describe regular penultimate accent, although production characteristics are unstudied. Recently, however, dialects from Guerrero state have been described with pitch-accent from historical coda breathy /h/ in nonfinal position [Guion and Amith, J. Acoust. Soc. Am. **117**, S2490 (2005)]. In the current study, the acoustic correlates of this new pitch-accent, as well as penultimate accent, are investigated. Speakers ($n=6$) from four dialects–two with innovated pitch-accent and two with only penultimate accent–were recorded. The production characteristics of fundamental frequency (F0), duration, and voice quality, as indexed by the difference between the first two harmonics (H1-H2), were investigated. Innovated pitch-accent was found to be cued solely by F0. Penultimate accent varied in its production by dialect. In dialects with only penultimate accent, production exhibited characteristics typical of stress-accent: Accented syllables had higher F0, longer durations, and more vocal effort (smaller H1-H2 differences) than unaccented syllables. Dialects with innovated pitch-accent had less stress-like penultimate accents, being cued primarily by F0, or with a relatively small durational difference, but no voice quality difference. This indicates that the innovation of a pitch-accent has changed the formerly stresslike penultimate accent into a more pitch-type accent.

2aSC35. Lexical stress without postlexical head-marking: Evidence from Tagalog. Victoria B. Anderson (Dept. of Linguist., Univ. of Hawaii, 1890 East-West Rd., Honolulu, HI 96822, vanderso@hawaii.edu)

Jun (2005) compares prosodic prominence types in 21 languages representing diverse families (including Australian, Chinese, Germanic, Muskogean, Romance, and Semitic, but none from Austronesian). Jun observes an implicational relationship between the prominence type a language uses at the lexical level (i.e., lexical tone, stress or pitch accent) and the prominence type it uses postlexically (i.e., head-marking or edge-marking). Fourteen languages in the sample employ lexical stress; all use postlexical head-marking (i.e., postlexical pitch accent on a phrases most prominent syllable), and four additionally mark a phrases' edge(s). No stress language in the sample employs postlexical edge-marking alone. However, Tagalog may be such a language. Tagalog uses lexical stress (LAyas "leave;" laYAS "carefree"), but arguably does not use head-marking. This study examined map-task data from four native speakers of Tagalog. Syllables bearing main stress in a phrase, and occurring between zero and three syllables from the end of the phrase, were analyzed for pitch, intensity, and duration. Results show that although syllables with primary stress have increased intensity, and duration, pitch prominences

do not reliably align with these syllables. Instead, low pitch reliably appears on the first syllable in an accentual phrase, while high pitch reliably appears on the last syllable.

2aSC36. Bengali intonational phonology. Sameer ud Dowla Khan (Dept. of Linguist., UC Los Angeles, 3125 Campbell Hall, UCLA, Los Angeles, CA 90095-1543)

Previous studies of Bengali intonation are primarily concerned with the mapping between syntax and intonation, the existence of accentual and phonological phrases, optionality in phrasing, and realization of focus. While these studies have been influential, they shy away from a full description of Bengali intonation, drawing from a small data set produced by a handful of speakers. At present, there is no comprehensive analysis of Bengali intonation that uses data collected from a wide range of speakers. The current study adopts the framework of intonational phonology (Pierrehumbert, 1980; Ladd, 1996) to develop such an analysis based on a large set of data from approximately 25 Bengali speakers. A pilot study measuring pitch and segment durations in 155 sentences suggests the existence of three pitch accent types (low, high, rising) and three levels of prosodic units whose right edges are marked by a boundary tone (intonation, intermediate, accentual). This suggests that Bengali differs from most other languages (Jun, 2005) in having both (1) contrastive pitch accents and (2) a small prosodic unit that is the domain of the pitch accent (i.e., the accentual phrase). Further analyses will illustrate the phonetic realizations of pitch accents and boundary tones in the language.

2aSC37. Prosodic correlates to meaning in infant-directed speech. Debora S. Herold (Dept. of Psych., Indiana Univ. Purdue Univ. at Indianapolis, 402 North Blackford St., Indianapolis, IN 46202, dssasso@iupui.edu), Laura L. Namy, and Lynne C. Nygaard (Emory Univ., Atlanta, GA 30322)

The present investigation examined the extent to which speakers produce reliable prosodic correlates to meaning across a range of semantic domains and whether parents use these kinds of cues to teach word meanings to children. In study 1, acoustic properties including fundamental frequency (F_0), F_0 variation, duration, and amplitude of infant-directed utterances were examined. Speakers were asked to produce question phrases in which novel words were used to convey one of two meanings from a set of antonym pairs (e.g., big/small, yummy/yucky). Herold (2006) has shown that adult listeners can identify word meanings from the prosodic information contained in these utterances. Acoustic analyses revealed that each word meaning had a unique acoustic signature, and that related word meanings such as *big* and *tall* and *small* and *short* had similar acoustic profiles. Study 2 explored whether naturalistic parental speech to young children recruits similar acoustic properties in the service of providing prosodic cues to meaning. Results demonstrated that mothers systematically varied prosodic cues as a function of word meaning when producing speech to their children. These findings suggest that speech directed to infants contains reliable prosodic markers to word meaning and that mothers may use prosodic cues to differentiate word meanings in word learning contexts.

2aSC38. Influence of voice prosody on preferential behaviors. Naoki Hasui and Yugo Takeuchi (1-39-18 Sumiyoshi, Hamamatsu-shi, Shizuoka, 432-8011 Japan, gs06044@s.inf.shizuoka.ac.jp)

In mother-infant interactions, the mother talks in a special voice that is characterized by elevated fundamental frequency (pitch) and exaggerated intonation contours, and the infant turns a face toward the voice and smiles naturally. This case suggests that humans react automatically to the special prosody without understanding of the lexical meaning in the word. We propose a hypothesis that adults react automatically as in the case of the infant by not only a difference of prosody but also the situation that one is in. This study aims to figure out the function of prosody on voice

cognition and utilize these findings for constructing the voice user interface systems with a low cognitive load for users. In this paper, using the virtual maze on a display, we gave the participants a task that they should escape as quickly as possible from the maze. In the situation where the participants choose the route on T-shaped intersection, we gave the participants two leading voices of different prosody, and observed the preferential behavior for the two voices. The experimental results indicate that the voice of the voice user interface systems synthesizing the high pitch or the high rate of speech is naturally accepted and can direct the users.

2aSC39. Utilization of acoustic cues in identifying speech intonation contrasts in cochlear implant recipients and listeners with normal hearing. Shu-Chen Peng (Dept. of Hearing and Speech Sci., Univ. of Maryland, College Park, MD 20742)

This study investigated the utilization of acoustic cues in identifying speech intonation contrasts in adult cochlear implant (CI) recipients and normal-hearing (NH) individuals listening to spectrally and temporally degraded stimuli. Fundamental frequency, intensity, and duration patterns of a disyllabic word, popcorn, were manipulated orthogonally, resulting in 360 resynthesized stimuli. In a two-alternative forced-choice task, CI and NH participants identified whether each stimulus was question-like or statement-like. Each NH listener also identified stimuli that were noise vocoded. Preliminary results from seven CI and four NH listeners indicated that (a) weighting of fundamental frequency in the identification of CI listeners was less pronounced than that of NH listeners, (b) unlike NH listeners who showed little reliance upon intensity or duration patterns in identifying unprocessed stimuli, CI users demonstrated systematic dependence upon intensity and duration cues in identifying the same set of stimuli, and (c) with noise-vocoded (spectrally and temporally degraded) stimuli, NH listeners' identification exhibited weighting patterns for inten-

sity and duration cues in a way similar to that of CI listeners. Implications for the processing of multidimensional acoustic cues for speech intonation in normal and electrical hearing will be discussed. [Work supported by NIDCD-R01DC04786.]

2aSC40. The voices of anger and disgust: Acoustic correlates in three languages. Marc D. Pell, Areej Alasseri (School of Commun. Sci. & Disord., McGill Univ., 1266 Pine Ave. West, Montreal, QC H3G 1A8, Canada, marc.pell@mcgill.ca), Sonja A. Kotz, and Silke Paulmann (Max Planck Inst. for Human Cognit. and Brain Sci., 04103 Leipzig, Germany)

Anger and disgust are believed to represent discrete human emotions with unique vocal signatures in spoken language. However, few investigations describe the acoustic dimensions of these vocal expressions, and some researchers have questioned whether disgust is robustly encoded in the vocal channel. This study sought to isolate acoustic parameters that differentiate utterances identified as sounding angry versus disgusted by listeners based on evidence from three separate languages: English, German, and Arabic. Two male and two female speakers of each language produced a list of pseudo-sentences (e.g., Suh fector egzullin tuh boshent) to convey a set of seven different emotions. The recordings were later judged by a group of native listeners to determine what emotional meaning was perceived from the prosodic features of each pseudo-utterance. Individual sentences identified systematically as conveying either anger or disgust (greater than $3\times$ chance target recognition) were then analyzed acoustically for various parameters of fundamental frequency, amplitude, and duration. Analyses compared which acoustic parameter(s) were dominant for identifying anger versus disgust in each language, and whether these patterns appeared to vary across languages, with implications for understanding the specificity and universality of these emotion expressions in the vocal channel.

WEDNESDAY MORNING, 29 NOVEMBER 2006   HONOLULU ROOM, 8:00 TO 11:55 A.M.

Session 2aSP

Signal Processing in Acoustics, Architectural Acoustics, Engineering Acoustics, Psychological and Physiological Acoustics: Spatial Sound Processing, Control, and Performance Evaluation

William L. Martens, Cochair
McGill Univ., Faculty of Music, 555 Sherbrooke St. W., Montreal, Quebec, H3A 1E3, Canada

Kazuhiro Iida, Cochair
Matsushita Electric Industrial Co., Ltd., AV Core Technology Development Ctr., 600 Saedo, Tsuzuki, Yokohama 224-8539, Japan

Chair's Introduction—8:00

Invited Papers

8:05

2aSP1. Median plane localization using a parametric model of the head-related transfer function based on spectral cues. Kazuhiro Iida, Motokuni Itoh (AV Core Technol. Development Ctr., Matsushita Electric Ind. Co., Ltd. 600 Saedo, Tsuzuki, Yokohama, 224-8539 Japan, iida.kazuhiro@jp.panasonic.com), Atsue Itagaki, and Masayuki Morimoto (Kobe Univ. Rokko, Kobe, 657-8501 Japan)

The present study has two purposes. One is to examine a simulation method for localizing a sound image. The authors propose a parametric HRTF model to simulate vertical sound localization. The parametric HRTF is recomposed only of the spectral peaks and notches extracted from the measured HRTF, and the spectral peaks and notches are expressed parametrically with center frequency, level, and sharpness. The other purpose is to clarify the contribution of each spectral peak and notch as a spectral cue. Localization tests were carried out in the upper median plane using the subjects' own measured HRTFs and the parametric HRTFs with various combinations of spectral peaks and notches. The results show that the parametric HRTF recomposed using the first and second notches

(N1 and N2) and the first peak (P1) provides almost the same localization accuracy as the measured HRTFs. Observations of the spectral peaks and notches indicate that N1 and N2 change remarkably as the source elevation changes, whereas P1 does not depend on the source elevation. In conclusion, N1 and N2 can be regarded as spectral cues, and the hearing system could utilize P1 as the reference information to analyze N1 and N2 in ear-input signals.

8:25

2aSP2. Evaluation of head-related transfer function on the opposite side to a sound source. Takanori Nishino (Ctr. for Information Media Studies, Nagoya Univ., Nagoya, Aichi, 464-8603 Japan), Takayuki Oda, and Kazuya Takeda (Nagoya Univ., Nagoya, Aichi, 464-8603 Japan)

A head-related transfer function (HRTF) is a complex acoustic transfer function because it includes the effect of reflection and refraction due to the head and ears. The HRTFs on the opposite side to a sound source have more complex characteristics and lower power than HRTFs that are on the same side as a sound source. Moreover, many previous studies about monaural sound localization showed that performances of sound localization on the opposite side were not precise. From these results, it can be assumed that the HRTFs on the opposite side are not necessary to use listener's own. In our experiments, several stimuli that convolved with the HRTFs were presented to evaluate individuality. The left ear HRTF was the subject's own, and the right ear HRTF was the subject's own and others. The differences between the subject's HRTF and others' HRTFs were evaluated using the spectral distortion score. The sound source azimuths were from 0 to 180 deg at intervals of 30 deg; the front was 0 deg, the left side was 90 deg, and the back was 180 deg. The results shows that the right ear HRTFs between 30 and 150 deg azimuth have low individualities.

8:45

2aSP3. Classification head-related transfer functions for an individualization procedure by a tournament-style listening-test. Yukio Iwaya and Yôiti Suzuki (Res. Inst. of Elec. Commun., Tohoku Univ., 2-1-1 Katahira, Aoba-ku, Sendai, 980-8577, Japan)

Individualization of head-related transfer functions (HRTFs) is necessary for high-accuracy sound image localization in three-dimensional spatial rendering by virtual auditory displays (VADs). However, measurement of HRTFs for each listener for all sound-source directions requires a special measuring apparatus and a long measurement time with a listener's physical load. The authors have therefore proposed an individualization method of HRTFs called Determination method of OptimuM Impulse-response by Sound Orientation (DOMISO), in which the best set of HRTFs is selected using a tournament-style listening test without any measurement. In DOMISO, our HRTF corpus, including more than 120 sets of measured HRTFs, is used. In this study, a classification method of these HRIRs is examined to select good seeds for the tournament. First, relationships among various parameters of HRTFs and individualization results by DOMISO are analyzed using multiple-correlation analysis. Second, using promising parameters, all sets of HRTFs in the corpus are classified as a structured tree. Third, using results of classification, listening tests by DOMISO procedure are performed using these HTRFs. Finally, a listening test of sound image localization is conducted with the individualized HRTFs to examine the effectiveness of the classification.

9:05

2aSP4. Development of the fast measurement system for the listener's own head-related impulse responses. Kimitoshi Fukudome (Dept. of Acoust. Design, Faculty of Design, Kyushu Univ., Shiobaru 4-9-1, Minami-ku, Fukuoka, 815 Japan, fukudome@design.kyushu-u.ac.jp), Yusuke Tashiro (Kyushu Inst. of Design, Minami-ku, Fukuoka, 815-8540 Japan), Kazuki Takenouchi, Toshiya Samejima, and Naoki Ono (Kyushu Univ., Minami-ku, Fukuoka, 815-8540 Japan)

Three-dimensional (3-D) spacial sound reproduction through headphones or earphones is best realized for subjects by the use of their own head-related impulse responses (HRIRs). Although HRIR measurements are so extremely time consuming, better externalized HRIRs for subjects must be collected in order to develop future 3-D applications. The fast measurement system for the listener's own HRIRs was constructed based on the continuous measurement method with a subject sitting on a servo-swiveled chair (SSC), and together with a loud speaker elevation arm controlled by a stepping motor. It can greatly reduce the subject burden in HRIR measurements. HRIRs for all azimuthal directions at 29 elevations can be measured in 1 h, constituting a reduction by a factor of 3 of the amount of time required as compared to systems composed of an SSC and a conventional manual elevation system. In the presentation, principles of the fast measurement system and the specifications of the loud speaker elevation arm are described. Performance evaluations are presented in terms of the measurement accuracies by the signal-to-deviation ratio and the spectral distortion, and further the reproducibility in repetitive measurements. [Work supported by the specific project research in the Faculty of Design, 2005.]

9:25

2aSP5. Capture and rendering of spatial sound over headphones. Ramani Duraiswami, Dmitry N. Zotkin, and Adam O'Donovan (Perceptual Interfaces and Reality Lab., Inst. for Adv. Comput. Studies, Univ. of Maryland, College Park, MD 20742)

A theory for capturing an audio scene and then rendering it remotely over headphones is developed. The method relies on capture of a sound field up to a certain order in terms of spherical wave functions. Then, the captured sound field is convolved with the head-related transfer function and rendered to provide presence in the auditory scene. The sound-field representation is then transmitted to a remote location for immediate rendering or stored for later use. A system that implements the capture using a spherical array is developed and tested. Head-related transfer functions are measured using the system described in [D.N. Zotkin *et al.*, J. Acoustic. Soc. Am. (to appear)]. The sound renderer, coupled with the head tracker, reconstructs the acoustic field using individualized head-related transfer functions to preserve the perceptual spatial structure of the audio scene. [Work partially supported by VA.]

2aSP6. Acoustic rendering of a virtual environment based on virtual microphone control and binaural room scanning. Jonas Braasch (School of Architecture, Rensselaer Polytechnic Inst., 110 Eighth St., Troy NY 12180), William L. Martens, and Wieslaw Woszczyk (McGill Univ., Montreal, QC H3A 1E3, Canada)

Binaural room scanning (BRS) is a convolution technique that utilizes a set of spatially indexed impulse responses that are selected in response to listener head movements during virtual acoustic rendering, such that the direction of sonic elements is automatically updated according to the angle determined by a head tracker. Since the room impulse responses have to be measured for only a few loudspeaker positions, a very successful application of BRS is the simulation of a control room. In order to create a flexible headphone-based virtual environment, it is proposed to simulate the input signals for the BRS system using virtual microphone control (ViMiC). ViMiC renders a virtual environment by simulating the output signals of virtual microphones that can be used to address a surround loudspeaker system. The advantages of this approach are twofold. First, the measured impulse responses of the BRS system ensure high spatial density in overall reflection patterns, avoiding the typical gaps in between the early reflections that occur when using solely mirror images or ray tracing. Second, the resulting imaging is closer to common audio engineering practice, where a pleasant but plausible environment is favored over a more purely realistic rendering.

10:05−10:20   Break

10:20

2aSP7. A comparison of listener loudspeaker preference ratings based on *in situ* versus auralized presentations of the loudspeakers. Sean Olive, Todd Welti (Harman Intl. Industries, Inc., 8500 Balboa Blvd., Northridge, CA 91329), and William Martens (McGill Univ., Montreal, PQ, H3A 1E3, Canada)

Auralization methods have several practical and methodological advantages for studying the perception of loudspeaker reproduction in rooms. For subjective measurements of loudspeakers, the auralization should be capable of eliciting the same sound quality ratings as those measured using an original acoustic presentation of a loudspeaker. An experiment was designed to test whether this is possible. Listeners gave preference ratings for both acoustic (*in situ*) and auralized double-blind presentations of four different loudspeakers, and the results were compared. The auralized presentations were generated from binaural room-scanned impulse responses of the loudspeakers convolved with the music test signals and presented over headphones equipped with a head-tracking device. For both *in situ* and auralized methods nine trained listeners gave preference ratings for four different loudspeakers using four different programs with one repeat (eight trials in total). The results show that the auralized and *in situ* presentations generally produced similar loudspeaker preference ratings.

Contributed Papers

10:40

2aSP8. Equalization of a one-dimensional sound field with a movable source. Xi Chen, Timothy W. Leishman, Scott D. Sommerfeldt, and Benjamin M. Faber (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

An experimental plane-wave tube with a movable side-branch loudspeaker was used to study simultaneous source and sound field equalization. The effects of equalization and inverse filtering were evaluated for different measured field quantities. Parametric equalization filters were designed based on acoustic pressure and total energy density at a point within the tube. An adaptive filter based on total energy density and the filtered-x algorithm was also implemented. The results of the equalization methods are compared and their advantages and disadvantages are explained.

10:55

2aSP9. Numerical calculation method of binaural signals suitable for auralization of a reflective sound field. Makoto Otani (Dept. of Intelligent Systems Design Eng., Toyama PRef. Univ., Imizu, Toyama 939-0398, Japan, otani@pu-toyama.ac.jp) and Shiro Ise (Kyoto Univ., Yoshida-Honmachi, Sakyo-ku, Kyoto City 606-8501, Japan)

Based on the Kirchhoff-Helmholtz integral equation, a numerical calculation method, which enables a wave-theory-based auralization of a sound field inside room, is proposed. Binaural signals are synthesized from pressures and particle velocities on the boundary surface and corresponding head-related transfer functions (HRTF). The proposed method can shorten the extra computation time that is required because of changes in numerical conditions caused by head movement and rotation. The boundary element method (BEM) was used to calculate surface values on sound-reflective objects. In addition, HRTFs for various source positions were calculated using a BEM-based HRTF calculation system [M. Otani and S. Ise, J. Acoust. Soc. Am. **119**, 2589–2598 (2006)]. Results obtained using the proposed method are compared with those of the conventional BEM, whose boundary consists of a head and a wall, and an image source method, which does not incorporate sound wave motion. Those comparisons reveal that the proposed method can accurately simulate binaural signals, including indirect sound from reflective objects. In addition, interactions between the head and wall are discussed, which are omitted in any auralization method using HRTFs to synthesize binaural signals and have not been investigated.

11:10

2aSP10. A new digital module for variable acoustics and wave field synthesis. Diemer de Vries, Jasper van Dorp Schuitman (Lab. of Acoust. Imaging and Sound Control, Delft Univ. of Technol., P.O. Box 5046, 2600 GA Delft, The Netherlands), and At van den Heuvel (Acoust. Control Systems B.V., 3886 ZG Garderen, The Netherlands)

In the 1980s, wave field synthesis (WFS) was proposed by Berkhout [J. Audio Eng. Soc., **36**, pp. 977–995 (1988)] as a new and powerful concept for spatial reproduction of sound fields without sweet-spot-limitations. Based on this concept, with implementation of simplifications not deteriorating the perceptual quality, a system for variable acoustics was developed: acoustic control system (ACS). The system has been widely applied as a solution for multipurpose halls. Until now, ACS processors have been built using analog electronic technology. Recently, however, a digital module has been designed that not only enables us to generate the ACS reflection patterns, but also to perform true WFS, e.g., for direct sound reinforcement with preservation of natural spatial properties. In the paper, design and specifications of the module will be discussed. Also, the potentials of integrated application of WFS and variable acoustics will be explained.

2aSP11. Multichannel high sampling rate 1-bit audio with a high definition video recorder system. Jun-ichi Fujimori (Sound and Acoust. Group, Sound Technol. Development Ctr., Yamaha Corp., 203 Matsunokijima, Iwata, Shizuoka, Japan, fujimori-j@beat.yamaha), Yasuhiro Oikawa, and Yoshio Yamasaki (Waseda Univ., Tokyo, Japan)

Ideally, as much audio, video, and metadata information as possible should be acquired to produce archives using the most advanced technology available at the time so that maximum information can be inherited. On the other hand, it is not easy to develop both audio technology and video technology simultaneously or even to integrate them. For example, because we specifically examine audio technology, it is desirable to use video products in the market to achieve the goal. In this sense, we propose the use of IEEE1394 equipment and synchronize their components for integration. We developed a high sampling rate 1-bit audio AD/DA converter that is equipped with IEEE1394 and recording and playback software running on Linux. The software can synchronize a 1-bit audio stream using MPEG-TS of HD video to extend its audio quality using IEEE1394 isochronous streaming with 125-Es resolution. This integration of audio and video is achieved without modifying existing video products. This paper describes the system and the 1-bit audio AD/DA converter (5.6 MHz sampling 8ch in 8ch out or 4.8 MHz sampling 16ch in 16ch out).

2aSP12. Audio noise reduction utilizing inharmonic-frequency analysis of generalized harmonic analysis (GHA). Teruo Muraoka (Res. Ctr. For Adv. Sci. and Technol., The Univ. of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo, 153-8904, Japan) and Thoru Ifukube (The Univ. of Tokyo, Tokyo, 153-8904, Japan)

Audio noise reduction is a desired technology for better telecommunications. Currently, the technology is based mostly upon spectral subtraction methods. However, its fundamental problem of musical noise generation remains unsolved. Musical noise is produced by mismatching of instantaneous phases of presampled noise and target noise to be canceled, which causes noise-breathing effects synchronous with the object's intensity variation. Additionally, unnecessary frequency components inherent to Fourier analysis to resynthesis process are mixed in the processed output. For this reason, spectral subtraction is inapplicable to high-fidelity audio. The authors have devoted attention to inharmonic frequency analysis of generalized harmonic analysis (GHA), which was originally proposed by Wiener. The analyses were intended to examine stochastic signals utilizing harmonic frequency analysis. Results clarified that any signal can be represented as an almost periodic function whose frequency components have an inharmonic relationship. In 1993, Dr. Hirata proposed inharmonic frequency analysis based upon successive matching with frequency component templates, which is applicable to high-fidelity audio signal processing. This method came to be called GHA, following Wiener's theory. The authors mostly attempted noise reduction in damaged SP records of historical value, retrieving almost noise-free musical sounds.

WEDNESDAY MORNING, 29 NOVEMBER 2006

MAUI ROOM, 7:50 A.M. TO 12:00 NOON

Session 2aUW

Underwater Acoustics and Acoustical Oceanography: Sediment Acoustic Processes: Quantifying the Effects of Sediment Properties on Dispersion, Attenuation, Reflection, Scattering, and Buried Target Detection I

Eric I. Thorsos, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6606

Masao Kimura, Cochair
Tokai Univ., Dept. of Geo-Environmental Technology, 3-20-1 Shimizu-Orido, Shizuoka 424-8610, Japan

Chair's Introduction—7:50

Invited Papers

7:55

2aUW1. The Sediment Acoustics Experiment in 2004 (SAX04). Eric I. Thorsos, Kevin L. Williams, Dajun Tang, Steven G. Kargl Darrell R. Jackson, and Brian T. Hefner (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, eit@apl.washington.edu)

A major sediment acoustics experiment (SAX04) was conducted near the coast of northwestern Florida, USA during September–November 2004. In addition to our laboratory, a number of other institutions participated, principally from the USA and Canada. SAX04 addressed sound penetration into, propagation within, and scattering from the shallow-water seafloor, as well as scattering from proud and buried targets. The sandy sediment at the experiment site has a critical angle of nearly 30 deg, convenient for study of acoustic penetration into sediments at grazing angles below the critical angle. A shallow-water depth of about 17 m allowed divers to be utilized in many aspects of the experimental work. In particular, divers deployed a 30-m rail system on the bottom for synthetic aperture sonar and other acoustic measurements. SAX99, conducted at a site nearby, showed the importance of scattering from ripples on high-frequency acoustic penetration at subcritical grazing angles. In part because of these findings, a new research initiative on sediment ripple formation, evolution, and decay was begun by ONR, and investigators from this program carried out studies at the SAX04 site. An overview will be given of the SAX04 measurement program. [Work supported by ONR.]

2aUW2. Observations of seafloor ripples and a time-dependent ripple geometry model. Peter Traykovski (Appl. Ocean Phys. Dept., Woods Hole Oceanogr. Inst., Woods Hole, MA, 02543)

Recent seafloor acoustic scattering experiments and theoretical developments have shown that the presence of ripples on the seafloor can increase the amount of subcritical angle acoustic penetration into the seafloor. A ripple observation and acoustic scattering experiment (SAX04) was recently conducted near Fort Walton Beach, Florida. During this experiment a sidescan sonar on an REMUS AUV was used to observe ripple geometry from the nearshore (5-m depth) to the shelf break (50-m depth). The ripples were found to increase in wavelength from 20 cm in the nearshore to 1.2 m at the shelf break. The offshore ripples were a product of Hurricane Ivan, which had passed the site 2 weeks prior. Time series of ripple geometry, measured with a tripod-mounted rotary sidescan sonar and a two-axis rotary pencil beam sonar, from a coarse sand site at the Martha's Vineyard Coastal Observatory, also revealed long-wavelength (0.6 to 1.2 m) relict ripples after wave events. Ripple geometry models that assume the ripples are in equilibrium with hydrodynamic forcing underpredict the wavelength of the relict ripples at both sites; thus a new model that allows for a time delay between ripple adjustment and forcing was developed. [Work supported by ONR.]

2aUW3. Biodegradation of wave-formed sand ripples during SAX04 (Sediment Acoustics Experiment 2004). Alex E. Hay (Dept. of Oceanogr., Dalhousie Univ., Halifax, NS B3H 4JI, Canada)

The gradual decay of 50- to 70-cm-wavelength sand ripples over a 20-day period of continuous *in situ* observations is presented. The ripples formed midway through the SAX04 experiment during Tropical Storm Matthew, and the observation period began immediately following this event. Bed elevation profiles and planform images of the ripple field were obtained at centimeter-scale resolution using rotary sonars, and at millimeter-scale resolution using a laser-video bed profiling system. Low-frequency currents, wave orbital velocities, and near-bed turbulence were monitored simultaneously with colocated instruments. The observed rates of change of the principal components of the bed elevation spectrum and of the spatially integrated backscatter intensity differences are presented and compared to predictions made with a diffusivity-based model of the ripple decay process. The model-data comparisons yield values for the horizontal sediment diffusivity which are consistent with previous measurements using particle-counting techniques.

2aUW4. Velocity dispersion in water-saturated granular marine sediments. Masao Kimura (School of Marine Sci. and Technol., Tokai Univ., 3-20-1 Orido, Shimizu-ku, Shizuoka, Shizuoka 424-8610, Japan, mkimura@scc.u-tokai.ac.jp)

The velocity dispersion in water-saturated granular marine sediments is an important subject that has remained unresolved in the community of sediment acoustics. Reported velocity dispersion results cannot be explained using the Biot model with the constant frame bulk modulus. The author showed that the values of the frame bulk modulus of water-saturated glass beads and beach sands are approximately 10^9 Pa at a frequency of 500 kHz and about ten times greater than those used typically in the community. Moreover, the author proposed a modified gap stiffness model incorporated into the Biot model (BIMGS model). Using this model, the author demonstrated that the frame bulk modulus has frequency dependence [M. Kimura, J. Acoust. Soc. Am. **1205** (2006)]. In this study, the BIMGS model is established using the measured results of the longitudinal wave velocities in glass beads saturated with alcohol-water mixed liquids. Then the velocity dispersion in glass beads and beach sands with different grain sizes are analyzed using the measured results of the longitudinal wave velocity at a much higher frequency and in the region of relaxation frequency. Finally, it is shown that the reported velocity dispersion can be explained using the BIMGS model.

Contributed Papers

2aUW5. Sound waves and shear waves in saturated unconsolidated sediments. Michael J. Buckingham (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0238)

A recent, continuum-mechanics theory of compressional and shear wave propagation in saturated unconsolidated sediments relies on grain-to-grain interactions as the sole source of internal stress in the porous medium. On the basis of a statistical argument, nonlinear microscopic processes occurring at grain boundaries are incorporated into the intrinsically linear grain-shearing theory. The result is two sets of algebraic dispersion relationships, one of which represents a compressional wave and the other a shear wave. The shear wave is predicted to exist even in the absence of an elastic "frame." For both types of wave, the theory predicts logarithmic dispersion and a near-linear dependence of attenuation on frequency. The grain-shearing dispersion relations, which are functions not only of frequency but also of the geophysical parameters of the sediment (porosity, bulk density, grain size, and overburden pressure), show excellent agreement with multiple data sets that have appeared in the literature over the past half century. Since the physical mechanisms behind the grain-shearing theory and Biot's classical theory of wave propagation in porous media are quite different, some comments will be offered in an attempt to clarify the circumstances in which the respective theoretical approaches might be valid. [Research supported by ONR.]

2aUW6. Frame virtual mass for unconsolidated granular media with slippage. Nicholas P. Chotiros and Marcia J. Isakson (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

A model of the grain contact physics in the presence of a propagating wave is extended to included losses due to slippage. Mechanical coupling between grains in a randomly packed unconsolidated granular medium is inherently stochastic. Grain-to-grain contact asymmetry leads to coupling of kinetic energy into orthogonal linear and rotational motion at the grain level, and partitioning of the energy. The result is an increase in inertia, which may be treated as an additional or virtual mass term for both shear and compressional waves. The virtual mass term can be evaluated, given the scintillation index of the compressional and shear contact stiffness, Poisson's ratio, and the average number of contact points at each grain. The effect of slippage at the grain-to-grain contact is examined. [Work supported by Office of Naval Research, Ocean Acoustics.]

9:45

2aUW7. A comparison of two observations of compressional wave velocity dispersion in marine sediments. Charles W. Holland (Appl. Res. Lab., The Penn State Univ., State College, PA), Altan Turgut (Naval Res. Lab., Washington, DC), Jan Dettmer, and Stan Dosso (Univ. of Victoria, Victoria, BC, Canada)

Compressional velocity dispersion exists in all marine sediments. If the dispersion is great enough, it may play a significant role in acoustic interaction with the seabed. On the other hand, if dispersion is weak, seabed models and databases can be substantially simpler. The ocean acoustics community is divided on this issue, in part because of the lack of observations. Here we compare two recent methods for measuring dispersion. One technique measures the broadband reflection coefficient (frequency dependence of the critical angle) and thus infers the velocity dispersion. The other technique measures time of flight in an *in situ* probe over a broadband of frequencies. Measurements on New Jersey Shelf (STRATAFORM area) are compared and implications for the frequency dependence of the compressional wave attenuation are discussed. [Work supported by the Office of Naval Research and NATO Undersea Research Centre.]

10:00–10:15   Break

10:15

2aUW8. Low-frequency laboratory measurements of sound speed in water-saturated granular sediments. Theodore F. ArgoIV (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029) and Preston S. Wilson (Univ. of Texas, Austin, TX 78712-0292)

Accurate knowledge of the acoustic behavior of water-saturated granular sediments is required for effective sonar operation in certain environments. Acoustic dispersion at frequencies below about 20 kHz has been the topic of much recent study and debate. While the understanding of real ocean-bottom sediments in their full complexity is the ultimate goal, laboratory measurements are significantly less difficult and subject to less uncertainty. This plays an important role in sediment acoustics since the difference between the predictions of competing propagation models is only a few percent. A new laboratory impedance tube apparatus, which operates in the 0.5–10-kHz frequency range, has been developed to address this issue. The device can be operated as a traditional acoustical impedance tube, in which water-borne sound waves interact with the water-saturated sediment. The device can also be operated as a driving-point impedance instrument, in which the mechanical input impedance of a column of water-saturated sediment is investigated. The effective-fluid sound speed is then inferred from the impedance measurements. Preliminary measurements of the sound speed in reconstituted natural sand sediments and artificial granular sediments obtained with the new apparatus will be presented and compared to existing theory. [Work supported by ONR.]

10:30

2aUW9. *In situ* acoustic scattering, attenuation, and dispersion measurements in marine sediments. Altan Turgut, Roger Gauss (Naval Res. Lab., Washington, DC 20375), Charles Holand (The Penn State Univ., State College, PA 16804), and Lisa Zurk (Portland State Univ., Portland, OR 97207)

The degree of acoustic velocity dispersion and nonlinear frequency dependency of attenuation in various types of sediments is investigated. A down-looking chirp sonar and a wideband acoustic probe system are used to measure compressional wave speed and attenuation within 2–150-kHz frequency band. Measurements indicate slight velocity dispersion and nonlinear frequency dependency of attenuation in silty and sandy sediments that can be effectively modeled by an extended Biot theory (Yamamoto and Turgut, J. Acoust. Soc. Am. **83**, 1744–1751 (1988)]. The degree of acoustic scattering due to volume inhomogeneities and surface roughness is also estimated from the incoherent components of the measured signals. Acoustic scattering results are favorably compared with those predicted by pseudospectral FDTD simulations that used colocated vibracore data as input. Robustness of the attenuation and dispersion measurements by using chirp sonar and acoustic probe systems is also verified by the FDTD simulations [Work supported by the ONR.]

10:45

2aUW10. Weak consolidation model for wave propagation in ocean sediments. James Berryman (Lawrence Berkeley Natl. Lab., 1 Cyclotron Rd., Berkeley, CA 94720)

Within the context of Biot's theory of poroelasticity, both effective density approximations and grain-to-grain elastic and inelastic effects are incoporated into a complete theory of acoustic propagation through ocean-bottom sands. The sedimentary material is assumed to be at best weakly consolidated. A variety of possible attenuation mechanisms are considered in order to explain ultrasonic data for both laboratory studies on glass beads saturated with water and silicon oil, as well as *in situ* data from the SAX99 experiment. An effective medium model is introduced to treat the weakly consolidated frame constants. Wave speeds and attenuations are explained to first order by the effective density approximation alone, but the remaining small discrepancies in velocity can be modeled successfully using the consolidation constants. Remaining attenuation discrepancies can then be fit using several methods including (a) increase of viscosity in fluid trapped between grains, (b) increase of viscosity due to high frequencies alone, and also (c) a new approach that specifically takes dynamic— but small amplitude—grain rearrangements into account. None of these mechanisms was taken into account in Biot's original low-frequency theory.

11:00

2aUW11. Frequency dependences of absorption and scattering components of attenuation for water-saturated sands. Keiichi Ohkawa (Tech. R&D Inst., JDA, 5-2 Ichigayahonmuracho Shinjuku-ku, Tokyo, Japan)

Frequency dependences of attenuation of sound waves in water-saturated marine sands are discussed. Measured attenuation includes a scattering component, and this makes it difficult to understand the frequency dependences of attenuation. By removing the scattering component from the measured attenuation, the Kramers-Kronig (K-K) relationship gives the correct sound speed dispersion. Consequently, the intrinsic attenuation extracted from the data collected during SAX99 experiments follows the Biot theory for overall frequencies. This suggests that the attenuation coefficient of sound waves for water-saturated sands is not linear in f. The sound speeds recalculated via the K-K relationship are also in excellent agreement with the Biot theory at high frequencies.

11:15

2aUW12. Dispersion and attenuation due to scattering from heterogeneities in the porosity of sand sediments. Brian T. Hefner, Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698), Joseph Calantoni, and Allen H. Reed (Naval Res. Lab., Stennis Space Ctr., MS 39529)

Close packed granular materials are inherently heterogeneous at small length scales due to the random packing of the grains. These heterogeneities should manifest themselves in the frame moduli and the porosity of the medium. For sound propagation through fluid-saturated granular materials, the small scale of the heterogeneities may lead to scattering from the fast compressional wave into the slow compressional wave, increasing the attenuation of sound in the medium. This mechanism may account for the increase in high-frequency attenuation relative to expectations based upon Biot theory observed in recent measurements in sand and glass-bead sediments. Previously, we presented the results of a perturbation theory of sound propagation that accounted for losses due to scattering from heterogenieties in the bulk frame modulus [B. T. Hefner *et al.*, J. Acoust. Soc. Am. **119**, 3447 (2006)] and showed that the losses due to this mechanism were not enough to account for the measured attenuations. The theory has since been expanded to account for heterogeneities in the porosity. This

theory relies on knowledge of the spatial statistics of the porosity and efforts are underway to measure these spatial variations in both simulations of glass beads and through CT scans of sand sediments. [Work supported by ONR.]

11:30

2aUW13. Measurements of the frequency dependence of sound speed and attenuation in a sandy sediment. Brian T. Hefner, Darrell R. Jackson, Kevin L. Williams, and Eric I. Thorsos (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698)

Among the systems deployed during the sediment acoustics experiments, SAX99 and SAX04, were a fixed buried hydrophone array and a diver-deployed mobile array. The buried hydrophone array was used to measure the sound penetration into the sediment from a mobile source tower over a frequency range of 11–50 kHz during SAX99. For very steep incident grazing angles well above the critical angle, the penetrating field is only weakly influenced by the surface roughness and can be used to determine the sound speed and attenuation in the sediment. For SAX04, the frequency range was extended such that the new range covered 2–50 kHz. The diver-deployed array consisted of two sources and two receivers, which could be inserted into the sediment to a depth of 10 cm to measure sound speed and attenuation from 100–260 kHz during SAX99. For SAX04, the frequency range was increased to 40–260 kHz such that the measurements overlapped those made using the buried array. The sound speeds measured during both experiments using these systems are consistent with the predictions of Biot theory while the attenuation follows a linear frequency dependence at high frequencies. [Work supported by ONR.]

11:45

2aUW14. Frequency dependence of sound speed and attenuation in marine sediments. William M. Carey (Boston Univ., Boston, MA 02215), Ji-Xun Zhou (Georgia Inst. of Technol., Atlanta, GA 30332), and Allan D. Pierce (Boston Univ., Boston, MA 02215)

Zhou and Zhang [J. Acoust. Soc. Am. (2005) (2006)] have reviewed diverse data for propagation in sandy sediments. Attenuation for frequencies below 1 kHz corresponds roughly to $\alpha = 0.34(f/f_o)^2$ dB/m with the reference frequency being 1 kHz. The best-fit exponent is slightly less than 2. Their review of results for frequencies above 10 kHz suggests a substantially different frequency dependence. They considered two explicit models, the Biot model as parametrized and extended by Stoll and the BICSQS model of Chotiros and Isakson [J. Acoust. Soc. Am. (2004)]. In either case the fit of parameters to match sound speed dispersion resulted in an unsatisfactory fit of attenuation versus frequency, and conversely. The present paper reexamines this and contends that, while Biot's formulation is intrinsically correct at low frequencies, Stoll's parametrization is noncausal: Chotiros' model more nearly captures the correct physics in that it is intrinsically causal and includes relaxation processes. A considered extension is to include a continuous smear of relaxation processes (rather than just two) within the basic context of the original Biot model. This is shown to be wholly consistent with Kramers-Kronig relations and to have sufficient flexibility to fit the overall data set.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006   LANAI ROOM, 1:00 TO 5:20 P.M.

Session 2pAA

Architectural Acoustics, Speech Communication, Musical Acoustics, Psychological and Physiological Acoustics, and Signal Processing in Acoustics: Special Session in Honor of Manfred Schroeder

Ning Xiang, Cochair
Rensselaer Polytechnic Inst., 110 8th St., Greene Bldg., Troy, NY 12180

Juergen Schroeter, Cochair
AT&T Labs., Research, 180 Park Ave., Bldg. 103 D163, Florham Park, NJ 07932-0971

Akira Omoto, Cochair
Kyushu Univ., Dept. of Acoustic Design, Shiobaru 4-9-1, Minami, Fukuoka 811-8540, Japan

Chair's Introduction—1:00

Invited Papers

1:05

2pAA1. Manfred R. Schroeder's life in acoustics: Phases, reflections, and other contributions of a "perfect number." Birger Kollmeier (Medizinische Physik, Universitaet Oldenburg, D-26111 Oldenburg, Germany)

Number theory such as the rules to compose a perfect number (which is the sum of all its divisors, e.g., 6, 28, 496, 8128) constitutes only a small part of the many fields of research Manfred R. Schroeder mastered in his scientific life. After finishing his Ph.D. with Erwin Meyer in Goettingen in 1954, Schroeder joined Bell Labs where he eventually became director of acoustics and mechanics research. In 1969 he was also appointed professor of physics and director of the Drittes Physikalisches Institut in Goettingen. He contributed to many seminal publications and inventions in the areas of room acoustics, speech coding, physiological and psychological acoustics, and computer graphics. The Schroeder phases describe the quadratic initial phase relation of the harmonics in a tone complex to produce a maximum flat (speech) waveform. The Schroeder diffusors are reflection gratings based on his number theoretic considerations that suppress the specular reflections while distributing scattered sound into all different directions. The current talk will give an overview of Manfred Schroeder's life in acoustics and his various contributions, including some stories about his unique character. A special focus will be placed on the Schroeder phases and their current use in hearing research.

<center>1:25</center>

2pAA2. Modification of time and frequency dimensions of speech. James L. Flanagan (Rutgers Univ., 26 Wolf Hill Dr., Warren, NJ 07059)

Vocal and auditory mechanisms impose constraints on speech signals. These constraints are central to techniques for coding and bandwidth compression, and hence to efficient transmission and storage. A key representation is the short-time Fourier transform, describing the signal in terms of two variables—time and frequency. This separation affords opportunity, within limits, to scale these information-descriptive parameters. This report recollects efforts (incentivized by bandwidth conservation) to modify speech in time and frequency. The ''Analytic Rooter'' [Schroeder *et al.*, Proc. IEEE **55**, 396–401 (1967)] divides bandwidth by a factor of 2, keeping temporal characacteristics intact. The ''Phase Vocoder'' [J. Flanagan and R. Golden, Bell Syst. Tech. J. **45**, 1493–1509 (1966)] allows nonrational manipulations in both time and frequency. And, ''Silence Detection,'' based upon automatic step-size adjustment in adaptive differential PCM coding [Flanagan *et al.*, IEEE Trans. Commun. **COM-30**, 336–345 (1982)], permits speed variations with constant bandwidth. Recordings illustrate the effects.

<center>1:45</center>

2pAA3. Early linear predictive speech coding at Bell Telephone Laboratories and Electrical Communication Laboratory. Fumitada Itakura (Meijo Univ., Shiogamaguchi 1-501, Tempaku-ku, Nagoya 468-8502, Japan)

At the 6th ICA in Tokyo (1968), two papers on the linear predictive coding of speech were presented by Manfred Schroeder and by me in the same session. The first paper was on the adaptive linear predictive waveform coding for medium band coding (time domain modeling) and the latter was on the LPC vocoder based on the autoregressive spectral estimation for narrow-band coding (frequency domain modeling). Although their approaches were very different in both concept and scope, there were dramatic common points: the covariance and autocorrelation methods were used to solve normal equation to determine the LPC parameters and also the pitch period was extracted to reproduce speech periodicity in both systems. Subsequently the multipulse (1982), stochastic (1984), and code (1985) excitation LP were proposed at AT&T BTL, whereas the transformation of LPC parameters to PARCOR (1969) and LSP (1975) and their efficient quantization were intensively studied at ECL of NTT. These efforts were eventually combined for practical implementations in the secure, cellular, and IP telephone in the 1980s. The brilliant ideas by Manfred Schroeder and his associates at BTL, in particular Bishnu Atal, have always been a helpful lighthouse to navigate the world-wide speech coding research in the right direction since 1980s.

<center>2:05</center>

2pAA4. Personal reflections on 20 years of research with Manfred R. Schroeder: From concert hall acoustics to code-excited linear prediction. Bishnu S. Atal (Dept. of Elec. Eng., Univ. of Washington, Seattle, WA 98185)

I joined Bell Laboratories, Murray Hill in 1961. Bell Labs then had an excellent environment for challenging researchers to find creative solutions to difficult problems in telecommunication. Coming from India, I was fortunate to have Manfred as my mentor. He and I collaborated on many interesting problems. Stereo sound systems in those days kept the sound confined between two speakers, and we wondered why we could not spread the sound beyond the speakers. We invented a system for creating signals at the human ears that removed this limitation, producing the beginning of surround sound. Vocoders could not synthesize natural-sounding speech, and Manfred always felt that the speech synthesis models used in vocoders were too rigid and inflexible. This kind of thinking led us to the path of adaptive predictive coding and we presented a paper on this topic at the first IEEE meeting on speech communication in Boston in 1967. We continued working together producing better and better speech quality on predictive coders. Our research led to the approach of code-excited linear prediction, which forms the basis of world-wide digital cellular standards.

<center>2:25</center>

2pAA5. The roles of models in speech communication research. Hiroya Fujisaki (Univ. of Tokyo, Tokyo, Japan)

The use of models is a key to development of science and technology. Only through modeling can one test various ideas, hypotheses, and theories against what one observes in nature. This paper first gives a systematic review of models that have been developed in the past on various aspects of speech communication, both between humans and between human and machine. These include speech production models, speech acoustics models, speech perception models, as well as language models and dialogue models. It will then describe briefly some of the models developed by the current author. Finally, it will discuss those aspects that have not been modeled so far, and present the author's prospects on how to model them. This paper is dedicated to Manfred Schroeder, whose distinguished contributions in many areas of acoustics were enlightening to a large number of fellow researchers.

<center>2:45</center>

2pAA6. Progress in electroacoustic transducer research. Gerhard M. Sessler (Darmstadt Univ. of Technol., Merckstrasse 25, 64283 Darmstadt, Germany, g.sessler@nt.tu-darmstadt.de)

Manfred Schroeder always had a vivid interest in the field of electroacoustic transducers. During his time as Director of Acoustics Research at Bell Laboratories, work on electret transducers was started by Jim West and the author in 1962 and the activities were extended over the years to electret materials, directional microphones, and microphone signal processing. Schroeder himself made some interesting contributions to this field, such as the design of directional microphones based on gradients of various orders with toroidal and unidirectional directivity patterns. Some of these devices were implemented with electret transducers. Work on directional microphones at Bell Labs continued as the author went to the Darmstadt University of Technology. There, studies on silicon transducers were initiated, with the first silicon condenser microphone described in 1983. Recently, a completely new and promising variety of piezoelectric transducers was implemented, consisting of charged cellular polymers. This is a new modification of the electret transducer principle developed in the 1960s at Bell Laboratories.

3:20

2pAA7. Application of digital signal processing to architectural acoustics in the 1960s. James E. West (Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

Digital signal processing techniques in architectural acoustics were first used to investigate the acoustical properties of Philharmonic Hall at Lincoln Center for the Performing Arts in New York City in the early 1960s. The management of Lincoln Center invited the Acoustics Research Department of Bell Labs, headed by Manfred Schroeder, to assist in an evaluation of the Hall because conventional measurement methods were unable to explain its subjective performance. Computer-generated specially shaped tone bursts were played and the sound pressures at different locations in the Hall were recorded and later converted into digital form for evaluation. Special filters were programmed to evaluate reverberation times based on different portions of the decay, and energy distribution in time, space, and frequency domains. Better insights into the relationships between subjective and physical measurements were made possible utilizing the power in digital signal processing.

3:40

2pAA8. New directions in room acoustics: Probabilities, computer simulation, number theory. Heinrich Kuttruff (RWTH Aachen Univ., D52066 Aachen, Germany, hku@akustik.rwth-aachen.de)

This paper highlights just a few of M. R. Schroeder's major contributions to room acoustics. In his first investigations Schroeder presented a new approach to the frequency response of rooms by recognizing its random character, at least in the range above a critical frequency limit (today known as the Schroeder frequency). Soon afterwards he began employing the then-new digital computer as a research tool. Together with B. S. Atal he simulated the sound transmission in rooms, and in 1961 both authors demonstrated what nowadays is known as auralization and cancellation of acoustical cross-talk between two loudspeakers. In 1967 they published probably the first paper on ray tracing. Furthermore, Schroeder contributed to measuring techniques, for instance by introducing the integrated impulse response method of recording sound decay. Another astounding achievement is Schroeder's invention of pseudo-random surface structures based upon number-theoretical schemes (Schroeder diffusers). Such devices may be used to scatter sound waves into all directions and hence to improve the acoustics of concert halls by creating lateral sound reflections. By his unconventional ideas and numerous innovations Manfred R. Schroeder has changed our understanding of sound transmission in rooms and initiated many new research activities.

4:00

2pAA9. Acoustic contribution to travel comfort in airplanes— insight delivered from room acoustics. Volker Mellert (Inst. f. Phys., Oldenburg Univ., 26111 Oldenburg, Germany)

Travel comfort in airplanes is affected by numerous parameters from which the acoustic environment provides a major contribution, as revealed in measurements of various symptoms of passenger and crew in long-haul flights. Tools known from room acoustics are useful to calculate and estimate an optimal acoustic design of the cabin interior, e.g., background noise levels must match the demand for privacy, or an appropriate STI ensures good public address and communication between the passenger and crew member. Many measures assume a statistical description of the acoustic field. But the strong frequency-dependant absorption leads to different Schroeder frequencies f_s, which indicate the transition from well-separated resonances to many overlapping normal modes. In consideration of the extended noise sources as boundary conditions it is useful to determine with f_s the regimes of analytical solutions of the wave field within the cabin from a more costly numerical (FEM/BEM) approach.

4:20

2pAA10. Are concert halls random number generators? Jean-Dominique Polack (UPMC, CNRS, MCC, LAM, 11 rue de Lourmel, F-75011 Paris, France)

In 1954, Manfred Schroeder proved that sound fields in rooms are statistical in nature. Based on his subsequent work, several models of sound fields were derived, both in the frequency and the time domain, that make it possible to predict most of the properties of sound fields in rooms, including the major room-acoustical indices. One of these models, derived by the author, is based on random number generation, another field to which Manfred Schroeder contributed. This model is presented, and its predictions compared to measurements by A. C. Gade and G. W. Siebein in US concert halls (see Gade's presentation at Sabine Centennial symposium in June 1994). Predictions are also compared with current theories of reverberation. However, the model is not able to predict the modal behavior below the Schroeder frequency yet. The solution lays in the so-called semiclassical approximation, which is briefly presented. The case of rectangular halls, where the semiclassical approach is no longer an approximation, but is exact, illustrates how to reconcile Monte Carlo approaches, ray theory, and mode distribution.

4:40

2pAA11. Schroeder's phase in psychoacoustics. Armin Kohlrausch (Philips Res. Labs. Eindhoven (WO02), HTC 36, NL-5656AE Eindhoven, The Netherlands, armin.kohlrausch@philips.com)

I want to address Manfred Schroeder's contributions to psychoacoustics with a special emphasis on the period 1980 to 1990, when I worked with him at the Third Physical Institute in Goettingen. Among the psychoacoustic topics Schroeder worked on were binaural hearing, often in relation to room acoustics and stereophony, and monaural phase perception. The latter topic was first addressed in an ASA conference abstract from 1959, in which Schroeder emphasized the role of the signal's peak factor for the perceived timbre. This theme was taken up in the famous mathematical paper in 1970, in which Schroeder demonstrated that a quadratic relation between frequency and phase in a (flat-spectrum) harmonic complex allows one to synthesize a periodic signal with a flat temporal envelope

2p WED. PM

and a quasilinear frequency modulation. Quite accidentally, it was found out in Bennett Smith's master thesis in 1984 that under specific circumstances, the input phase of the stimulus and the phase characteristic of a fixed point on the basilar membrane can nearly compensate each other. This interpretation, published in 1986 in J. Acoust. Soc. Am. with Schroeder as senior author, initiated a great number of perceptual and physiological studies on the phase characteristics of the inner ear.

5:00

2pAA12. Sequences and their consequences. Michael Vorlaender (Inst. fuer Technische Akustik, RWTH Aachen Univ., D-52056 Aachen, Germany)

Number theory is an inspiration for application in science and engineering. In acoustics, correlation sequences for measurement signals and for diffusors became famous, mainly due to innovative ideas of Professor Manfred Schroeder. [Number Theory in Science and Communication, 3rd ed. (Springer, Berlin, 1999).] Maximum-length sequences were implemented in many real-time analyzers for measuring impulse responses. Their autocorrelation properties are almost ideal and they are widely in use today. Benefits and limitations due to hardware constraints are briefly discussed. Another particular interesting sequence is the Fibonacci sequence. It can be created from a simple loop algorithm as a binary sequence. [H., Schmidt, ''Eigenschaften und Realisierung von Fibonacci-Folgen.,'' Proc. DAGA88, Braunschweig, p. 621]. The ratio of zeros and ones is converging to the golden ratio, 1.6180 . . . . Its spectrum consists of peaks at frequencies and with amplitudes corresponding to the golden ratio, too. When playing such sequences with certain clock frequency, a strange tonal sound can be heard. In psychoacoustic listening tests the pitch perception and pitch strength were investigated. Moreover, intervals and melodies were played by using a midi interface and new melodic and rhythmic pieces of Fibonacci music were created and checked in listening tests.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006 KOHALA/KONA ROOM, 1:00 TO 4:55 P.M.

Session 2pAB

Animal Bioacoustics: Fish Bioacoustics: A Tribute to Arthur Myrberg II

Richard R. Fay, Cochair
Loyola Univ. Chicago, Parmly Hearing Inst., 6525 N. Sheridan Rd., Chicago, IL 60626

Arthur N. Popper, Cochair
Univ. of Maryland, Dept. of Biology, College Park, MD 20742

Chair's Introduction—1:00

Invited Papers

1:05

2pAB1. Recent discoveries and continuing mysteries in the auditory systems of bony fishes. Peggy L. Edds Walton (Parmly Hearing Inst., Loyola Univ. Chicago, Chicago, IL 60626, peddswa@phi.luc.edu)

Recent research has revealed aspects of the auditory system that reflect the diversity of bony fishes as a group. This presentation will review some of the more exciting finds. For example, the long-standing classification of the hair-cell orientation patterns in saccules may require revision given the diversity of hair-cell orientations documented in closely related fishes recently. Hearing studies have shown that, although auditory specialist fishes (otophysines, with auditory accessories for detection of the pressure component of a sound wave) hear higher frequencies than auditory generalists (those that respond primarily to the particle motion component of a sound wave), work on the toadfish has shown that sensitivity at their best frequencies is not as different as once believed. However, a new kind of auditory specialist has emerged within the herrings. Some herring species respond to ultrasonic frequencies (¿20 kHz), apparently via a specialized region of the utricle. The utricle may be the rising star of the fish auditory periphery as physiological and anatomical-projection studies in nonherring species indicate that the endorgan may serve both the vestibular and the auditory systems. Emerging insights into auditory processing among fishes will be discussed. [Research supported by the NIH.]

1:25

2pAB2. Recent advances in hearing and sound communication by fishes. Friedrich Ladich (Dept. of Neurobiology and Behavior, Univ. of Vienna, 1090 Vienna, Althanstr. 14, Austria)

Acoustic displays have been observed in representatives of 30 or more families of bony fishes. Thanks to Art Myrberg's ground-breaking work, we have insights into the multitude of functions of acoustic signalling in fishes including mate choice, individual and species recognition, reduction of territory intrusions, and interception. Recent studies, many of which have arisen from Art's work, focus on ontogenetic and seasonal changes, female vocalizations, and effects of ambient and anthropogenic noise in hearing and sound communication. In croaking gouramis (labyrinth fishes), for example, ontogenetic development of acoustic communication depends on changes in sound parameters, increases in sound levels, and auditory sensitivities. The importance of seasonal changes in hearing was recently uncovered in toadfish and cyprinids. Generally male fish display acoustically to attract females, but the opposite may

occur. Only female croaking gouramis vocalize during mating. Ecoacoustical investigations reveal that hearing specialists (cyprinids) are well adapted to the low ambient noise levels of standing freshwaters. They are heavily masked in flowing waters, unlike generalists, which are merely affected by ambient noise. Gaussian and ship noise diminishes a fish's ability to detect sounds and resolve temporal patterns, indicating that human-made noise affects acoustic orientation and communication in fishes. [Work supported by FWF.]

1:45

2pAB3. Seasonal variation in the frequency response of the ear: Implications for communication. Joseph Sisneros (Dept. of Psych., Univ. of Washington, Seattle, WA 98195)

Seasonal variation in vocal-acoustic behavior is well known among nonmammalian vertebrates; however, studies of the neural mechanisms underlying such seasonal variation have focused mainly on the anatomical rather than the neurophysiological correlates of seasonal reproductive periodicity. In this talk, I present the neurophysiological evidence for seasonal variation in the frequency response of the ear in the plainfin midshipman fish (*Porichthys notatus*). Vocal communication is essential to the reproductive success of the seasonally breeding midshipman fish, which migrates during the late spring and summer from deep ocean sites into the shallow intertidal zone to breed. Female midshipman fish use the auditory sense to detect and locate males that are singing from their nests. The seasonal onset of male mate calling in the midshipman fish coincides with a dramatic increase in the range of frequency sensitivity of the female's saccule, the midshipman's main organ of hearing, thus leading to increased encoding of the male's mate call. I will present evidence for a steroid-dependent mechanism that leads to this form of auditory plasticity and discuss why this plasticity of peripheral auditory frequency sensitivity may represent an adaptation to enhance the acquisition of auditory information needed for mate identification and localization during the breeding season.

2:05

2pAB4. Directional hearing in sharks and other fishes. Anthony Hawkins (Loughine, Kincraig, Blairs, Aberdeen AB12 5YT, UK)

Art Myrberg and his co-workers were among the first biologists to obtain clear evidence that predatory sharks are able to locate sound sources at long distances. Previous experiments had shown that elasmobranches were relatively insensitive to sounds, and it had been suggested that fish were incapable of determining the direction of a sound source in the far field. It was subsequently shown that fish could discriminate between sounds from spatially separated sources in the horizontal and median vertical planes. They could even distinguish between identical sound sources at differing distances. It has been suggested that fish determine direction using particle motion detectors, the otolith organs of the inner ear. An array of vector detectors is capable of detecting the axis of propagation, but as they are inherently bidirectional there remains a problem in detecting the actual direction of the source. This paper considers the proposal that the ambiguities are resolved by phase comparison between the vector detectors and other detectors responding to the sound pressure incident upon the gas-filled swimbladde.

Contributed Papers

2:25

2pAB5. Sound production in two loricariid catfishes. Amanda L. Webb and Michael E. Smith (Dept. of Biol., Western Kentucky Univ., Bowling Green, KY 42101)

Many families of catfishes produce sounds via pectoral spine stridulation and/or swim bladder compression using sonic muscles linked to the swimbladder. Sound production capabilities in the catfish family Loricariidae has not yet been examined. We recorded sounds produced by two species of loricariid catfish, *Otocinclus affinis* and *Pterygoplichthys gibbiceps*. Both species produced short, broadband clicks via pectoral spine stridulation. Mean fundamental and dominant frequencies for *O. affinis* and *P. gibbiceps* were approximately 4000 and 8000, and 600 and 1200 Hz, respectively. Durations of sound pulses were generally shorter in *O. affinis* (2–12 ms) compared to *P. gibbiceps* (20–100 ms). Scanning electron microscopy was used to examine the dorsal process of the base of the pectoral spines of each species. Sound is produced as the ridges of the pectoral spine base slides past the channel of the pectoral girdle. Distances between dorsal process ridges of *O. affinis* and *P. gibbiceps* were approximately 50 and 160 microns, respectively, with the ratio of this distance to

the fish total length being comparable between species. Relationships between spine interridge distance, pulse duration, and spectrum of the sounds produced will be discussed. [We thank John Andersland for SEM assistance.]

2:40

2pAB6. Sound production and hearing ability in the Hawaiian sergeant fish. Karen P. Maruska, Kelly S. Boyle, Laura R. Dewan, and Timothy C. Tricas (Dept. of Zoology, Univ. of Hawaii and Hawaii Inst. of Marine Biol., 46-007 Lilipuna Rd., Kaneohe, HI 96744, maruska@hawaii.edu)

Sounds provide important signals for inter- and intraspecific communication in fishes, but few studies examine behavioral contexts of sound production and hearing ability within a single species. This study characterized the sounds and their behavioral contexts in a wild population of Hawaiian sergeant damselfishes (*Abudefduf abdominalis*), and compared

their features to hearing sensitivity measured by both auditory-evoked potentials (AEP) and single-cell recordings in the auditory midbrain. The sergeant fish produces low intensity ($\sim$85–115 dB rms re: 1 μPa), low-frequency ($\sim$100–400 Hz) sounds during nest preparation, courtship an agonistic interactions. Both AEP and single-cell recordings show that fish are most sensitive to these frequencies (80–300 Hz) with best frequency at 100 Hz. However, auditory thresholds to tonal stimuli were approximately 20 dB lower for single-cell recordings than those measured by AEP experiments. Further, midbrain auditory neurons were on average more sensitive to playbacks of conspecific sounds than to pure tones. These experiments show that the hearing ability of this damselfish closely matches the intensity and spectral characteristics of sounds produced in the wild. Multidimensional studies that incorporate analyses of and responses to biologically relevant natural sounds are necessary to understand acoustic processing and the evolution of acoustic communication in fishes.

2:55–3:10   Break

3:10

2pAB7. Sound production mechanism in carapid fish: First example with a slow muscle. Eric Parmentier, Jean Paul Lagardere, Pierre Vanderwalle, Jean-Baptiste Braquegnier (Lab de Morphologie Fonctionnnelle et Evolutive, Univ of Liege, Belgium), and Michael L. Fine (Virginia Commonwealth Univ., Richmond, VA 23284-2012)

Fish sonic swimbladder muscles are the fastest muscles in vertebrates and have fibers with numerous biochemical and structural adaptations for speed. Carapid fishes produce sounds with a complex swimbladder mechanism including skeletal components and extrinsic sonic muscle fibers with an exceptional helical myofibrillar structure. To study this system we stimulated the sonic muscles, described their insertion and action, and generated sounds by slowly pulling the sonic muscles. We find the sonic muscles contract slowly, pulling the anterior bladder and thereby stretching a thin fenestra. Sound is generated when the tension trips a release system that causes the fenestra to snap back to its resting position. The sound frequency does not correspond to the calculated resonant frequency of the bladder, and we hypothesize that it is determined by the snapping fenestra interacting with an overlying bony swimbladder plate. To our knowledge this tension release mechanism is unique in animal sound generation.

3:25

2pAB8. Sound communication by the forceps fish, *Forcipiger flavissimus* (Chaetodontidae). Kelly S. Boyle and Timothy C. Tricas (Dept. of Zoology, Univ. of Hawaii at Manoa, 2538 McCarthy Mall, Edmondson 152, Honolulu, HI 96822)

The forceps fish, *Forcipiger flavissimus* (family Chaetodontidae), lacks the anterior swim bladder horns and laterophysic connection that are proposed to improve processing of acoustic signals in the genus *Chaetodon*. Despite the lack of hearing specializations, laboratory experiments using hydrophones synchronized to video revealed the forceps fish produce at least three sounds: (1) a pulse (21-ms duration, peak frequency=318 Hz) associated with rapid cranial elevation, (2) a low-frequency pulse (9-ms duration, peak frequency=41 Hz) associated with anal fin retraction and elevation, and (3) a high-frequency broadband click (3-ms duration, peak frequency=7924 Hz) associated with tail slap behavior. Auditory evoked potential experiments demonstrate that hearing thresholds for this species are 10–21 dB higher than *Chaetodon* in the 200–600-Hz band. High-speed kinematic analyses show that motor movements associated with some acoustic behaviors are different than those in *Chaetodon*, thus the presence of sound production may be the ancestral state for the family but the mechanisms are divergent. Field observations on Hawaiian reefs reveal that forceps fish occur in pairs and trios that may be haremic, and use acoustic communication among group members.[Work supported by NSF IBN0137883 to TT.]

3:40

2pAB9. Acoustico-lateralis communication in coral reef butterflyfishes. Timothy C. Tricas and Kelly S. Boyle (Dept. of Zoology and Hawaii Inst. of Marine Biol., Univ. of Hawaii at Manoa, Honolulu, HI 96822, tricas@hawaii.edu)

Butterflyfishes are conspicuous members of coral reefs and well known for their pairing behavior and visual displays during social interactions. Members of the genus *Chaetodon* have a unique arrangement of the anterior swim bladder horns that connect with the lateral line (the laterophysic connection) and also project towards the inner ear, but functions for this putative acoustico-lateralis adaptation were not known. Field experiments demonstrate that the monogamous multiband butterflyfish, *C. multicinctus*, produces hydrodynamic pulses and sounds during territorial defense, and pulse trains that may be an alert call to the mate. In field choice tests, individuals could distinguish mates from nonmates and produced different calls to each, thus demonstrating the existence of mate recognition and context-specific sound production. Laboratory experiments show that sound production differs among sexes, hydrodynamic pulses are of low frequency ($<$100 Hz), and sounds have peak energy from 100–500 Hz. Auditory evoked potential experiments demonstrate that the auditory system has best sensitivity from 200–600 Hz and that the swim bladder horns provide a 10–20-dB gain within this band. These social sounds produced by *Chaetodon* can provide relevant stimuli to the auditory, vestibular, lateral line and possibly the laterophysic systems.

3:55

2pAB10. Ultrasound hearing and response in alosine herrings: A short history of the pure and applied research to date. Carl R. Schilt (LGL Ltd., P.O. Box 416, North Bonneville, WA 98639)

In the early 1990s it was reported that downstream-migrating American shad (*Alosa sapidissima*) avoided a sonar-sampling beam with fundamental frequency over 150 kHz. Since then there has been a considerable body of laboratory and field work to determine which fishes have such hearing capacity, what the hearing limits are, what the hearing mechanisms may be, what might have led to selection for such capacity, and the utility of such stimuli to control the local distributions of fishes with such ultrasonic hearing. So far it seems that ultrasonic click aversion is limited to one subfamily of primarily anadromous herrings (family *Clupeidae*; subfamily *Alosinae*). The capacity to hear very high clicks and the propensity to swim away from the stimulus direction has led to experiments in shad protection, especially in hydropower and other industrial environments, which have been promising. This talk will review the history of both the pure and applied science on alosine herring ultrasound response and suggest possible research and application strategies for the future.

4:10

2pAB11. Continental shelf fish choruses in Australian waters. Robert D. McCauley (CMST Curtin Univ., G.P.O. Box U 1987, Perth 6845 Australia) and Douglas H. Cato (Defence Sci. & Technol. Org., NSW)

Almost everywhere the authors have sampled ambient noise on Australia's continental shelves, in tropical and temperate waters, evening choruses have been recorded. Chorus maximum frequency varies from a few hundred Hz to a few kHz and may result from different species, but the consistency of the timing and similar chorus nature suggests that different source species are following similar behavior. Individual chorus signals are comprised of single pops or clicks of $<$200-ms duration, similar to those described from fish. We hypothesize that these chorus sources are nocturnally active fishes that mainly forage in the water column at night. In the Perth Canyon, Western Australia, shelf break choruses centered at 2 kHz have been measured since 1992. Here each evening the chorus begins shortly after dusk, reaches a maximum around 2 hours post dusk, drops to a lower level, rises again predawn, and then disappears during the day. Canyon net sampling through the water column has only recovered fish of the family Myctophidae as potential chorus sources. Long sea-noise time series records from the Canyon indicate significant seasonal trends in chorus levels and differences between years, possibly indicating that chorus activity reflects secondary productivity.

2pAB12. Transmission of air gun array signals in Australian waters.
Robert D. McCauley, Susan J. Rennie (Curtin Univ. of Technol., G.P.O. Box U 1987, Perth 6845, Australia), and John R. Hughes (Santos Ltd., Adelaide SA 5000, Australia)

Received air gun array signal parameters from 15 seismic surveys in Australian waters have been analyzed. Measurements include sources from 1115 to 4900 cui (18.3 to 80.3 L), ranges of hundreds of meters to thousands of kilometers along travel paths of 15 m to full ocean depth. For any array there may be large (± 9 dB) difference in received level between consecutive signals, especially if two 3-D arrays are operated alternatively. All arrays demonstrate horizontal directivity, being typically omnidirectional below nominally 50 Hz but focusing higher frequency energy about their longer axis. Three-dimensional arrays tend to be square and 2-D arrays rectangular with the long axis across track; thus 2-D arrays are more focused. At a nominal range, differences of up to 40 dB were found when comparing small arrays operating in shallow water with large arrays in deeper water. For example, a 2500-cui array transmitting in 70 m water had no waterborne energy at 20-km range whereas a 4900-cui array operating on the shelf slope was detectable via a deep sound channel path at 2000-km. This large variation in received levels under different scenarios implies similar differences in environmental implications.

2pAB13. Distribution of static pressures on the skin surface of parallely swimming fish.
Kazuhiko Anraku, Tatsuro Matsuoka, and Kenji Watanabe (Faculty of Fisheries, Kagoshima Univ., 4-50-20 Shimoarata, Kagoshima 890-0056, Japan, anraku@fish.kagoshima-u.ac.jp)

Hydrodynamic static pressures on the body surface of fish generated by conspecifics parallely swimming were measured in a flume tank, and they were compared with those generated by changing the position of two individuals under different flow velocities. Two dead common carp (BL = 20 cm), fixed in 20% of formalin, were used. Pressure was measured by a manometer connected to the ends of three aluminum pipes (inner diameter = 3 mm), which were inserted perpendicularly to the body axis from right to left in the control fish. The openings of the pipes were located on head, trunk, and caudal regions, and each was on the infraorbital or trunk canal lateral lines. The target fish was moved on the horizontal XY plane using a traverse device. Although pressures changed depending on the regions measured, changes of pressure according to position and current velocity were relatively low in the caudal region. The absolute values of the pressure increased as flow velocity increased, and decreased with an increase in the distance between individuals. Relative positions when the maximum pressure changes were observed at each flow velocity did not change, indicating that pressures were largest when the maximum body width of the conspecific is closest.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006   WAIALUA ROOM, 1:00 TO 5:15 P.M.

Session 2pAO

Acoustical Oceanography and Underwater Acoustics: Acoustic Scattering by Aquatic Organisms IV

Kenneth G. Foote, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Masahiko Furusawa, Cochair
Tokyo Univ. of Marine Science and Technology, 4-5-7 Konan, Minato, Tokyo 108-8477, Japan

Chair's Introduction—1:00

Contributed Papers

2pAO1. Swimbladders under pressure: Anatomical and acoustic responses by walleye pollock.
John K. Horne, David Barbee (School of Aquatic and Fishery Sci., Univ. of Washingtons Box 355020, Seattle, WA 98195), Kouchi Sawada, Koki Abe, Kazuhiro Sadayasu, and Yoshimi Takao (Natl. Res. Inst. of Fisheries Eng., Kamisu-city, Ibaraki 314-0408, Japan)

Pressure influences echo intensities of swimbladdered fish through changes in swimbladder volumes and surface areas. Volume reduction is expected to correspond to Boyle's law (volume $\propto$ pressure^{-1}) but the resulting deformation in swimbladder surface area will largely determine target strength at geometric scattering frequencies. Dorsal and lateral radiographs of juvenile walleye pollock (*Theragra chalcogramma*) in a pressure chamber were used to image swimbladders from ambient to a maximum of 5-atm pressure. As pressure increased, dorsal swimbladder surface areas decreased at a constant rate among the three individuals. Swimbladder volume reduction rates were similar among individuals but less than that predicted by Boyle's law. Compression of swimbladders occurred dorsal-ventral, anterior-posterior, and laterally. Resulting swimbladder shapes became more spindlelike as pressure increased. KRM predicted target strengths at 38 and 120 kHz decreased more rapidly at 3 atm above ambient than at lower pressures. [Work supported by ONR and NRIFE.]

2pAO2. Modeling of the backscattering by swimbladder-bearing fish.
Dezhang Chu, Timothy K. Stanton (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), J. Michael Jech (NOAA/Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543), and D. Benjamin Reeder (Naval Postgraduate School, Monterey, CA 93943)

A composite scattering model has been developed to describe the backscattering by swimbladder-bearing fish. This model includes the influence of elongation of the swim-bladder not only on the resonance frequency, as do most of the existing scattering models, but also on the quality factor Q and the overall scattering level. In addition, within the region covering the transition from resonance scattering to geometric scattering, the effect of asymmetric cross section of the swimbladder has also been included to represent a more realistic shape of fish swimbladder. The model was compared with other scattering models, such as Kirchhoff ray mode (KRM) and Fourier matching method (FMM). Model predictions were then compared with field data collected with a new broadband towed system. For a single-species Atlantic herring school (*Clupea harengus*) with a measured length distribution, the predicted resonance scattering including resonance peak location and overall level matched the broadband acoustic data reasonably well. Some modeling issues will also be discussed.

1:35

2pAO3. Target strength and behavior of swimming juvenile southern bluefin tuna in a cage estimated by echo-trace analysis. Koki Abe, Yoshimi Takao, Kazuhiro Sadayasu, Kouichi Sawada, and Kazuo Amakasu (Nat. Res. Inst. Fish. Eng., 7620-7 Hasaki, Kamisu, Ibaraki, 314-0408, Japan)

Echo-trace analysis (ETA) was applied to echoes from juvenile southern bluefin tuna (*Thunnus maccoyii*, SBT) swimming in a cage. The experiments were conducted offshore Esperance in Western Australia, using a split-beam echosounder operating at 70 kHz. The dimensions of a cylindrical cage were 7 m in diameter, 20 m in depth, and 8.5 cm mesh size of a net. A transducer was set at the center of a cage. The samples of SBT were caught by trolling near the study site. The age classes were 1 or 2 years old that were ranging 45 to 82 cm in fork length. Only one SBT was put into a cage at a time in order to obtain single echoes from one fish. As the swimbladder of juvenile SBT was not gas-filled in this stage, its target strength (TS) was much lower than TS of a same-size fish with gas-filled swimbladder. Because of low TS values, there were some difficulties to obtain high quality echo-traces continuously. So, preselections of high signal-to-noise and noninterference echoes from the echogram were needed before applying ETA. Swimming velocities and TS patterns in dorsal aspect were estimated by ETA.

1:50

2pAO4. An advanced 3-D scattering model for squid. Benjamin A. Jones, Andone C. Lavery, and Timothy K. Stanton (WHOI, Woods Hole, MA 02543-1050)

A three-dimensional, acoustic scattering model for squid has been developed that incorporates high-resolution, spiral computerized tomography (SCT) scans of squid morphology. This model numerically implements the distorted wave Born approximation (DWBA) for an inhomogeneous body using a combined ray trace and volume integration approach. SCT scans of squid, taken in air, revealed detailed morphology of the animal including the interior cavity of the mantle. The model accounts for these features with volume elements that are small (less than 1/20 the wavelength) for the frequencies of interest (0–120 kHz). Target strength predictions are compared with published data from live, freely swimming squid. Although only two materials, mantle tissue and seawater, are used to represent squid in this study, our model can be easily generalized to include many distinct, weakly scattering regions within a volume. Ranges of validity with respect to material properties and numerical considerations are explored using benchmark computations with such simpler geometries as spherical and cylindrical shells.

2:05

2pAO5. Target strength estimation of the large jellyfish *Nemopilema nomurai*. Miyuki Hirose (Grad. School of Fisheries Sci., Hokkaido Univ., 3-1-1, Minato Hakodate Hokkaido, 041-8611, Japan), Tohru Mukai, Kohji Iida (Hokkaido Univ., Hakodate Hokkaido, 041-8611, Japan), and Doojin Hwang (Chonnam Nat. Univ., Dundeok-dong, Yeosu Jeollanam-do 550-749, Korea)

The jellyfish *Nemopilema nomurai*, which reaches up to 2 m in diameter and 200 kg in weight, has occurred in large numbers in the Sea of Japan during the last several years and has had a negative effect on coastal fisheries in this region. Data on the abundance and distribution of jellyfish

are needed to forecast when and where they will occur in coastal areas. Acoustic techniques are commonly used to study the distribution and abundance of fish and zooplankton because these techniques can survey large areas relatively quickly. However, before such surveys can be conducted, the acoustic characteristics of the target species must be known. In this study, the density of and speed-of-sound in live jellyfish were measured to clarify their acoustic characteristics using a theoretical scattering model. Jellyfish density was measured using the dual-density method, while the speed of sound was measured using the time-of-flight method. The acoustic characteristics were estimated with the distorted-wave Born approximation (DWBA) model using these material properties and the shape of free-swimming jellyfish. The results demonstrate the feasibility of investigating the abundance and distribution of jellyfish using acoustic methods. [Work supported by JSPS and KOSEF Core University Program on Fisheries Science (FiSCUP).]

2:20

2pAO6. Predictions of acoustic scattering by a fish school. Charles F. Gaumond and Ralph N. Baer (Naval Res. Lab., Washington, DC 20375)

A theory originally developed by Richard Love [J. Acoust. Soc. Am. **64**, 571–580 (1978)] is used to model scattering by a school of fish. In order to incorporate propagation effects, this model has been integrated into a modified version of the RAM acoustic propagation model developed by Michael Collins. Running the integrated model over a sequence of frequencies results in predictions of the time and frequency response of broadband signals scattered by fish schools. Using an ensemble of fish school realizations lets us develop features that describe the response of a fish school to the acoustic signals. These features in turn allow separation of returns from a target from those from interfering schools. [Work supported by the Office of Naval Research.]

2:35

2pAO7. Using acoustic backscatter to determine the size distribution of channel catfish in a commercial pond. John Heffington, James Chambers (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, 1 Coliseum Dr., Univ., MS 38677), David Heikes, Steve Pomerleau, and Eric Stone (The Univ. of Arkansas at Pine Bluff, Pine Bluff, AR 71601)

With increasing seafood demand, aquaculture is poised to become a major growth industry in the United States. In particular, channel catfish represents an approximately $400 million industry. Pond management strategies, however, are compromised by high fish densities with large size distributions that result from years of continuous culture without a total harvest. Recent work has been presented on the Aquascanner SONAR system, which evaluates the total biomass of the pond but is less effective in determining the size distribution. To address this shortcoming, research colleagues at the University of Arkansas Pine Bluff are using a subsample seine net. Their method involves pulling a trawl across the pond to obtain a sample of the fish and manually weighing the fish in a time-consuming process to provide the size distribution. The work presented here attempts to use acoustic backscatter to determine the size distribution of the fish in lieu of the manual weight measurements. After being seined, the fish are allowed to swim back into the pond through a restrictive pipe and are pinged acoustically to determine their target strength and thus size. The experimental design of the system and results will be discussed. [Work Supported by the USDA.]

Contributed Poster Papers

All posters will be on display from 1:00 p.m. to 5:00 p.m. Contributors of all papers will be at their posters from 4:25 p.m. to 5:15 p.m.

2pAO8. Swimbladder morphology and target strength of myctophid fish of the northwestern Pacific. Hiroki Yasuma, Kazushi Miyashita (Field Sci. Ctr. for the Northern Biosphere, Hokkaido Univ., 3-1-1, Minato, Hakodate, Hokkaido, 041-8611 Japan), Kouichi Sawada (Natl. Res. Inst. of Fisheries Eng., Hasaki, Kashima, Ibaraki, 314-0421 Japan), and Ichiro Aoki (Univ. of Tokyo, Bunkyo, Tokyo, Japan)

Swimbladder shapes of fish from 14 myctophid species captured around Japan were measured using an x-ray method to estimate the theoretical target strength. Using those x-ray images, the fish were classified into bladder species and nonbladder species according to the presence of the gas. Five species had no swimbladder in any body size range examined. Others contained bladder and nonbladder fishes in relation to growth. The growth of swimbladder length was not proportional to that of body length (allometric growth) in some bladder fishes. Based on morphological parameters of the body and swimbladder, the relationships between body length and target strength from 38 to 200 kHz were estimated using theoretical acoustical backscattering models. Estimated values of target strength showed greater than 10-dB differences between bladder and nonbladder fishes, even when the fish were of equivalent size. Because of the allometric growth of the swimbladder, values of the target strength were not proportional to the square of the body length in most bladder species. In addition, larger fishes (over 60 mm) of Myctophum asperum showed regressive growth of the swimbladder (or target strength), suggesting that the use of the 20 log relationship is inappropriate for bladder species of myctophid.

2pAO9. *In situ* **measurement of target strength, tilt angle, and swimming speed of Boreopacific gonate squid (*Gonatopsis borealis*).** Kouichi Sawada, Hideyuki Takahashi, Koki Abe, and Yoshimi Takao (Natl. Res. Inst. of Fisheries Eng., Fisheries Res. Agency, 7620-7 Hasaki, Kamisu, Ibaraki, 314-0408, Japan)

Target strength (TS), mantle length (ML), swimming speed, and tilt angle of Boreopacific gonate squid *Gonatopsis borealis* were measured using an acoustical-optical composite system (J-QUEST) in the North Pacific Ocean. The J-QUEST consists of a quantitative echosounder operating at 70 kHz and a high-sensitivity stereo video camera system installed in a housing resistant to pressure at the depth of 300 m. In the daytime, Boreopacific gonate squids were seen at the J-QUEST depth of 230 m by the stereo video camera. Swift vertical movements were also observed on the echogram at the same time. Swimming speeds and tilt angles were estimated by tracking echoes from the same squid. Estimated average swimming speed was over 1 m/s and tilt angles were almost vertical. Average TS calculated from squidlike echoes was -54.1 dB and average ML calculated from the stereo video images was 23.9 cm with standard deviation of 5.9 cm. Standardized TS by the squared ML was -81.9 dB. In the nighttime, a Boreopacific gonate squid (estimated ML 37 cm) at the J-QUEST depth of 25 m was observed preying on one Japanese anchovy, *Engraulis japonicus*, in an aggregation with estimated average length of 12.3 cm.

2pAO10. Target strength of *Sargassum* plant with vesicles containing gas. Kenji Minami (Hokkaido Univ., 3-1-1 Minato-cho, Hakodate, Hokkaido, 041-8611, Japan), Akira Hamano, Takeshi Nakamura (Natl. Fisheries Univ., Shimonoseki, Yamaguchi 759-6595, Japan), Yoshimi Takao (Fisheries Res. Agency, Kamisu, Ibaraki, 314-0408, Japan), and Kazushi Miyashita (Hokkaido Univ., Hakodate, Hokkaido, 041-8611, Japan)

Information about the target strength (TS) of a vesicle is important to estimate the acoustic scattering of *Sargassum* plants. Experimental TS (TSe) of a vesicle was obtained at 70 and 200 kHz using quantitative echo sounders in a freshwater tank ($10 \times 10 \times 15$ m). The TSe of ten vesicles examined here were -56.88–51.68 dB at 70 kHz and -60.30–-50.90 dB at 200 kHz. The TSe's of small vesicles were 2.56–-3.70 dB higher at 70 kHz than at 200 kHz, suggesting that these small vesicles were affected by the airbladder resonance. Theoretical TS (TSt) was estimated by airbladder−scattering models to confirm this effect. The gas radii of 418 vesicles (containing the experimented ten vesicles) were measured using a soft x−ray system and applied to the theoretical models. Their radii were 0.52–-6.00 mm; calculated TSt matched with TSe. According to the theoretical model, the vesicles were not affected by the resonance when argument ka is $ka > 1$ (where k is the wave number and a is the radius). However, the vesicles were affected by the resonance when $ka < 1$. That fact indicates that small vesicles (a is 1.22–-2.38 mm) at 70 kHz were affected by the resonance.

2pAO11. Target strength of a mesopelagic fish *Maurolicus japonicus* at 38 and 120 kHz. Tadanori Fujino, Hiroki Yasuma, Kazushi Miyashita (Hokkaido Univ., 3-1-1, Minato-cho, Hakodate, Hokkaido, 041-8611, Japan, fnori@fish.hokudai.ac.jp), Kazuhiro Sadayasu, and Koki Abe (Fisheries Res. Agency, Hasaki, Kamisu, Ibaraki, 314-0408, Japan)

Target strength (TS) of *Maurolicus japonicus* was examined at 38 and 120 kHz using experimental measurements (at 38 kHz) and theoretical swimbladder scattering models. Experimental measurements showed that the average TS for dominant size (32.1 to 34.8 mm standard length) were -60.5 to -56.7 dB, and the effect of the fish orientation on TS was slight. These results coincided with the model estimation, proving that swimbladder models are valid to estimate the TS of *M. japonicus*. The slight effect of the fish orientation on TS was presumed due to the oval shaped swimbladder with high aspect ratio, which is relatively little to the wavelength. This is a merit because tilt angle is usually a major source of TS variation. However, a little swimbladder would cause acoustic resonance at low frequencies, which leads to insatiability on TS according to depth change. Using the resonant model, relationships of the TS and standard length were estimated for depth 50, 100, 150, 200, and 250 m at 38 and 120 kHz, respectively. While TS at 120 kHz showed little change according to depth, TS at 38 kHz changed with depth and showed larger TS than 120 kHz for each size.

2pAO12. Target strength pattern measurement of krill (*Euphausia pacifica*) and measuring limits in a small tank. Kouichi Sawada (Natl. Res. Inst. of Fisheries Eng., Fisheries Res. Agency, 7620-7, Hasaki, Kamisu, Ibaraki, 314-0408, Japan), Tohru Mukai (Hokkaido Univ., Hakodate, 041-8611, Japan), Kazuo Amakasu, Koki Abe, and Ken Ishii (Fisheries Res. Agency, Kamisu, Ibaraki, 314-0408, Japan)

Euphausia pacifica is an extremely important species for Japanfs northeast coastal fisheries and a key species for ecosystems in the northwest Pacific Ocean. Target strength (TS) of krill information have been required for acoustic surveys; they constitute a scale factor that is used to convert acoustic estimates of abundance. A precise TS pattern measurement system was constructed in a small tank. Target strength patterns of a target are measurable in the angle range of 60 deg at 1-deg steps. First, TS values of the thinnest fishing line with 53-mm diameter were measured as it was planned for use in tethering krill. TS measurement limits were also confirmed by measuring echo signals and noise signals. Maximum TS of the fishing line was -95 dB and the measurement limit was -98 dB at the frequency of 200 kHz with 10 dB of signal-to-noise ratio (SNR). Second, we confirmed the validity of the target suspension method. The thinnest fishing line was penetrated into the krill body to tether it between two vertical fishing lines. Finally, TS patterns of 12 live krill were measured and compared with theoretical calculations using distorted wave Born approximation. Both agreed well.

3:10–3:25   Break

Contributed Papers

3:25

2pAO13. Estimation of swimming behavior and backscattering pattern of Antarctic krill using echo-trace-analysis. Kazuo Amakasu (Natl. Res. Inst. of Fisheries Eng., Fisheries Res. Agency, 7620-7 Hasaki, Kamisu, Ibaraki, 314-0408, Japan, soundproc@yahoo.co.jp), Masahiko Furusawa (Tokyo Univ. of Marine Sci. and Technol., Minato-ku, Tokyo, 108-8477, Japan), and Sakiko Ando (Ocean High Technol. Inst. Inc., Nakano-ku, Tokyo, 164-0012, Japan)

Echo-trace-analysis was applied to single echo-traces of Antarctic krill to estimate *in situ* swimming orientation, swimming speed, and backscattering pattern. Acoustic data were collected with a split-beam echosounder operating at 70 kHz on 11 January 2005 during the Southern Ocean survey by the training vessel *Umitaka-maru* of Tokyo University of Marine Science and Technology. While the acoustic data were collecting, the ship speed was 2 knots. A pulsewidth and a pulse-repetition-period of the echosounder were 0.6 ms and 0.27 s, respectively. Krill aggregations were observed at the depth range 15–30 m from the hull-mounted transducer and at upper and lower edges of them some single echo-traces were visually recognized on echograms. In this work, these echo traces were analyzed. As preliminary results, the average swimming orientation was 4.0 deg with a standard deviation of 14.2 deg and the average swimming speed was 0.33 m/s with a standard deviation of 0.16 m/s. The average swimming orientation was close to the orientation observed with a video plankton recorder by Lawson *et al.* [J. Acoust. Soc. Am. **119**(1), 232–242 (2006)]. The estimated backscattering patterns showed peaks at incident angles around -10 to 10 deg and were similar to the patterns predicted by theoretical models.

3:40

2pAO14. Development and application of multifrequency acoustic techniques to the study of Antarctic krill distribution. Gareth L. Lawson, Peter H. Wiebe, Timothy K. Stanton, and Carin J. Ashjian (Woods Hole Oceanog. Inst., Woods Hole, MA 02543)

Multifrequency acoustic techniques were refined and applied in a study of the distribution of Antarctic krill. Acoustic, video, and net data were collected in a continental shelf region west of the Antarctic Peninsula in the falls and winters of 2001 and 2002. Improved parametrization of a theoretical model of krill target strength was achieved through direct measurement of all model parameters. Krill visual acuity was used to determine a threshold level of volume backscattering strength for identifying krill aggregations. Differences in mean volume backscattering strength at 120 and 43 kHz further served to distinguish krill from other sources of scattering. Inversions were performed of mean volume backscattering strength (43, 120, 200, and 420 kHz) in these acoustically identified ag-

gregations to estimate the weighted mean length and density of krill. Inversion results were comparable to net samples for estimated length, but acoustic estimates of krill density exceeded those from nets by one to two orders of magnitude, likely due to avoidance and differences in the volumes sampled by the two systems. Application of these methods to survey data demonstrated strong spatial and temporal variability in the horizontal and vertical distribution of krill aggregations, consistently associated with food availability, temperature, and predators.

3:55

2pAO15. Acoustic studies investigating the distribution and abundance of krill in critical blue whale habitat in the Monterey Bay submarine canyon. Louise P. McGarry, Charles H. Greene (Cornell Univ., Snee Hall, Ithaca, NY 14853), Peter H. Wiebe (Woods Hole Oceanog. Inst., Woods Hole, MA 02543), and Donald A. Croll (Univ. of California, 100 Shaffer Rd., Santa Cruz, CA 95060)

Blue whales, the largest animal on the planet, are frequently found in concentrations at the ocean surface in areas over thick sound-scattering layers of krill. In this study, the dive behavior of a blue whale was investigated in the context of zooplankton acoustic backscatter data concurrently collected in the Monterey Bay submarine canyon. The whale dive behavior revealed that the whale spent greater than 50 percent of its time in the region of the most intense backscatter, suggesting threshold foraging behavior. In order to describe the prey field within the canyon, additional day and night acoustic surveys with MOCNESS net tows in the canyon and along its perimeter were collected. Here, a three-dimensional characterization of the prey field is developed using the acoustic data. Further work will be undertaken to estimate the relative biomass of the zooplankton taxa generating the acoustic backscatter within and without the intense sound-scattering layer, and to model the acoustic backscatter that would be expected to be generated from the zooplankton specimens collected: the forward problem. [This work is supported by the ONR.]

4:10

2pAO16. Multiple-angle acoustic scattering and classification of zooplankton. Paul L. D. Roberts and Jules S. Jaffe (Marine Physical Lab, Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

The use of multiple-angle acoustic scattering to discriminate between two ecologically important classes of zooplankton' copepods and euphausiids, is proposed. The distorted wave Born approximation is used to model the multiple-angle scatter from a linear array of transducers designed to be practical for field deployment. Using a large set of noisy training and test data generated from uniformly random length and three-dimensional orientation distributions from each class of scatterers, the

performance of classification algorithms in several feature spaces is evaluated. The results show a marked improvement in classification performance as additional angles, and larger angular separations, are included in the classifier. Interestingly, even in the case of uniformly random three-dimensional scatterer orientation, low classification error (~5%) can be obtained. These results hold promise for substantially improving the classification of fluidlike zooplankton *in situ* using multiple angle scatter obtained with a simple collection geometry.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006   KAHUKU ROOM, 1:00 TO 3:15 P.M.

Session 2pBBa

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Shock Wave Therapy II

S. H. R. Hosseni, Cochair

Tohoku Univ., Inst. of Fluid Science, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

Robin O. Cleveland, Cochair

Boston Univ., Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Contributed Papers

1:00

2pBBa1. Radiation force imparted on a kidney stone by a Doppler-mode diagnostic pulse. Oleg A. Sapozhnikov (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA, 98105), Leonid A. Trusov (Moscow State Univ., Moscow 119992, Russia), Alexander I. Gromov (Mandryka 2nd Central Clinical Military Hospital, Moscow 107014, Russia), Neil R. Owen, Michael R. Bailey, and Lawrence A. Crum (Univ. of Washington, Seattle, WA, 98105)

Detection of kidney stones and estimation of their sizes is an important part of the lithotripsy treatment. Fluoroscopy is often used to target stones, but not every stone is radio-opaque and, in addition, fluoroscopy produces ionizing radiation. Acoustic waves offer an alternative way to visualize stones. The acoustic impedance of kidney stones typically differs significantly from that of surrounding tissue. A useful consequence of the impedance mismatch is the possibility to target stones with diagnostic mode ultrasound. Another consequence is that radiation force pushes the stone. Stone displacement may be responsible for the twinkling artifact that has been observed by several authors in color Doppler mode of ultrasound imaging. This effect can be used to detect not only renal and ureteral stones, but also calcifications in other organs (e.g., breast). In this paper we model the radiation force associated with the Doppler diagnostic pulse. The problem is divided into three parts: (1) acoustic scattering; it is solved in finite differences; (2) radiation force calculation; (3) stone velocity estimation supposing the stone sits in soft tissue. [Work supported by NIH DK43881, DK55674, NSBRI SMS00402, and RFBR.]

1:15

2pBBa2. Numerical solutions of the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation satisfying the Rankine-Hugoniot condition. Gianmarco Pinton and Gregg Trahey (Duke Univ., 136 Hudson Hall, Box 90281, Durham, NC 27708, gfp@duke.edu)

A numerical method that solves the KZK equation and satisfies the Rankine-Hugoniot relation for shock-wave propagation is described and characterized. By comparison with a known planar solution, it is demonstrated that current numerical methods in both the time and frequency domains predict a shock front that is stationary relative to the propagation phase and that the proposed method predicts the correct speed. These methods are then compared in the context of shock-wave lithotripsy and high intensity focused ultrasound. At the focus, axisymmetric shock-wave lithotripter simulations show that the Rankine-Hugoniot method predicts peak positive pressures that are 20% smaller, peak negative pressures that are 10% larger, and a full width at half maximum that is 45% larger.

High-intensity focused ultrasound simulations for an axisymmetric transducer in water have even larger variations in peak positive pressures and intensity but smaller variations in peak negative pressure. Simulations for a rectangular transducer in tissue, where absorption is more prominent, exhibit smaller but significant variations in peak pressures and intensity. Satisfying the Rankine-Hugoniot condition plays an important role in numerical simulations of the KZK equation. [This work was supported by NIH Grants R01-HL075485 and 1R01-CA114093-02.]

1:30

2pBBa3. Shock-wave energy deflection due to the presence of bone. Thomas Matula, Juan Tu, Michael Bailey (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Kirsten Fagnan, and Randy LeVeque (Univ. of Washington, Seattle, WA 98105)

The physical interaction of shock waves with musculoskeletal tissues is inherently different from other high-pressure acoustic therapies. Whereas high-intensity focused ultrasound and lithotripsy focus their energy in regions of soft tissue, musculoskeletal shock-wave therapy (SWT) involves focusing shock waves (SWs) near or at bones. The presence of bones will cause reflection, refraction, and diffraction of acoustic energy. In our analysis of cavitation generated during clinical SWT treatment, we noticed that the cavitation was occurring away from the SW focus. We confirmed this by performing an *in-vitro* experiment on a bone embedded in gel. We also quantified the deflection angle as a function position near a bone in water: A talus was manufactured using rapid prototyping. It was placed near the focus of an electrohydraulic SW device. To visualize the pressure field, a high-speed camera imaged the cavitation field generated around the focus. As the bone was moved closer to the focus, the cavitation field (and thus, the pressure field) deflected away from the bone. The deflection angle was measured as a function of relative distance between the bone and focus. Numerical simulations were performed to model the deflection of energy away from the bone.

1:45

2pBBa4. A sensitive, broadband polyvinylidine difluoride (PVDF) hydrophone for accurate characterization of shock waves. Adam D. Maxwell, Brian E. MacConaghy (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA), Oleg A. Sapozhnikov (Moscow State Univ., Moscow, 119992, Russia), and Michael R. Bailey (Univ. of Washington, Seattle, WA)

Few hydrophones are capable of measuring high-intensity fields such as shock waves accurately. One of the more reliable is the fiber-optic probe hydrophone. However, this system is expensive and insensitive. We

2p WED. PM

created a new PVDF hydrophone and compared it with a fiber-optic system. The hydrophone consisted of a 25-μm-thick PVDF membrane with a 0.5-mm active element and a preamplifier, which were each held in separate attached polycarbonate housings. The amplifier had adjustable gain and could account for membrane resonance to flatten the frequency response. A model of the frequency response for the system was developed, which agreed well with the measured response. Shock waves were measured in two Dornier HM-3 clones and an electromagnetic lithotripter. Measurements were also recorded using a 2-MHz focused piezoceramic source and a broadband PVDF source. Shock-wave measurements closely matched those recorded by the fiber-optic hydrophone and calculations made with a KZK-type model. Very little damage to the membrane was found after applying several thousand shock waves. This new membrane hydrophone is robust and sufficiently accurate to measure high-intensity fields, while greatly reducing cost, increasing sensitivity, and simplifying measurements of shock waves. [Work supported by NIH DK43881, DK55674, and NSBRI SMS00402.]

2:00

2pBBa5. Calculation and measurement of acoustic scatter to assess fragmentation in shock wave lithotripsy. Neil R. Owen (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Oleg A. Sapoznikov, Michael R. Bailey, and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

Shock wave lithotripsy (SWL) is currently conducted with little feedback on whether kidney stones are breaking. To determine if fragmentation could be assessed, acoustic scatter from intact and fractured stone models was calculated numerically and measured *in vitro*. Acoustic scatter from the stones, which were modeled with glass spheres, was calculated numerically using a linear elastic model, initialized with known elastic constants, and propagated from the stone model surface using the Helmholtz-Kirchhoff integral. Experimentally, shock waves were generated with a research lithotripter and scatter was measured with a broadband, spherically focused receiver. Calculated and measured results agreed well in the time domain. In frequency, power spectra were integrated to find energy and showed that scatter from the fractured stone model had higher energy in specific frequency bands that were related to the reverberation period. High-speed photography indicated that cavitation did not adversely affect the analysis of scatter. In this work it was possible to distinguish between the intact and fractured stone models. This method could be applied to stones that fragment gradually under the application of shock waves and potentially be used to estimate fragment size, and therefore the endpoint of therapy. [Work supported by NSBRI SMS00402 and NIH DK43881.]

2:15

2pBBa6. Bubble responses to lithotripsy shock waves. Wayne Kreider, Michael R. Bailey (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Oleg A. Sapozhnikov (Moscow State Univ., Moscow 119992, Russia), and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

The responses of bubbles subjected to a lithotripsy shock wave have been investigated numerically and experimentally to elucidate the role of heat and mass transfer in the underlying dynamics of strongly excited bubbles. Single spherical bubbles were modeled as gas-vapor bubbles by accounting for liquid compressibility, heat transfer, vapor transport, vapor trapping by noncondensable gases, diffusion of noncondensable gases, and heating of the liquid at the bubble wall. For shock-wave excitations, the model predicts bubble growth and collapse, followed by rebounds whose durations are significantly affected by vapor trapping. To experimentally test these predictions, bubble rebound durations were measured using passive cavitation detectors, while high-speed photographs were captured to evaluate the local cavitation field and to estimate radius-time curves for individual bubbles. Data were acquired for bubbles in water with varying temperature and dissolved gas content. Measurements verify that vapor trapping is an important mechanism that is sensitive to both temperature and dissolved gas content. While this work focuses primarily on individual

bubbles, some bubble cloud effects were observed. Analysis with a simple multibubble model provides noteworthy insights. [Work supported by NIH T32-EB001650, NIH DK43881, and NSBRI SMS00402.]

2:30

2pBBa7. Millisecond initiation of boiling by high-intensity focused ultrasound in tissue-mimicking phantoms. Michael S. Canney, Michael R. Bailey (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698), Vera A. Khokhlova (Moscow State Univ., Moscow 119992, Russia), and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105-6698)

Nonlinear propagation effects leading to shock formation at the focus of high-intensity focused ultrasound (HIFU) treatments can accelerate heating and cause rapid boiling in tissue. Boiling can be utilized for targeting the treatment with B-mode ultrasound and should be taken into account when planning the treatment, because bubbles reflect ultrasound and thereby displace and distort the lesion shape. In these experiments, an HIFU transducer of 2-MHz frequency, 4-cm aperture, and 4.5-cm focal length was used to investigate heating effects from shock formation in tissue-mimicking phantoms. The time required to attain 100 °C at the focus was calculated with weak shock theory from the peak amplitudes calculated with a KZK-type model, and time to boiling was measured by high-speed video and a 20–MHz passive cavitation detector (PCD) for different values of phantom absorption (both lower than tissue absorption) and HIFU power (100–200 W). Boiling was observed in 3 ms at the highest power level used by the observation of visible bubbles and by a significant change in the PCD time signal and spectrum. [Work supported by NIH DK43881, NSBRI SMS00402, and RFBR.]

2:45

2pBBa8. Characterization of the YDME FEP-BY02 tumor therapy system. Bryan Cunitz (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105), Stuart Mitchell, Yak-Nam Wang, Barbrina Dunmire, and Peter Kaczkowski (Univ. of Washington, Seattle, WA 98105)

The YDME FEP-BY02 is an extracorporeal tumor therapy device that has been in clinical use in China for several years. In the USA, this device is currently under investigation for the treatment of pancreatic cancer. Prior to being submitted for an application for an Investigation Device Exemption (IDE) with the FDA, the device will be required to undergo a thorough characterization of all systems involved in patient treatment. The purpose of this research is to characterize the FEP-BY02 and obtain data for an IDE application to the FDA for the treatment of pancreatic cancer using the FEP-BY02. A series of acoustical characterization, *in vitro*, and *ex vivo* tests are being conducted to validate the company's preclinical and clinical experience in China. Comparisons between three methods (radiation force balance, fiber-optic hydrophone, and scattering hydrophone) for measuring/determining focal intensities will be used to assess the validity of the measurement techniques and evaluate acoustic output over the functional input range of the FEP-BY02. Transparent gels impregnated with thermally sensitive markers will be used to visualize lesion formation and evaluate lesion characteristics with respect to input parameters. Thermocouples implanted in tissue mimicking thermal gels and *ex vivo* tissue will be use to evaluate the thermal response with respect to input parameters. Preliminary data indicate that the therapeutic ultrasound treatment system forms consistent lesions at the desired target.

3:00

2pBBa9. Acoustic virtual laboratory: A modeling tool for high-intensity focused ultrasound applications. Francesco P. Curra, Steven G. Kargl, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA)

The field of high-intensity focused ultrasound (HIFU) is emerging with strong potential and broad medical applications. Characterized by its ability to penetrate at depth inside the body without harming intervening tissue, therapeutic ultrasound has posed the basis for a new array of nonin-

vasive therapies. However, the inherent complexity of biological media and the nonlinear nature of ultrasound propagation at HIFU regimes make optimization and control of the therapy still a challenging task. In this respect, an accurate modeling tool, the Acoustic Virtual Laboratory (AVL), will be presented for solving multidimensional HIFU problems in complex geometries that can greatly assists in predicting HIFU applications effects and, therefore, help in the optimization and control of the treatment. AVL consists of 2-D, 2.5-D (cylindrical symmetry), and 3-D coupled solutions for acoustic and elastic wave propagation in heterogeneous, lossy, bubbly, and porous media with the bioheat equation for temperature estimation. It includes linear and nonlinear wave propagation, arbitrary frequency power law for attenuation, and can account for multiple reflections and backscattered fields. Sample results for tumor ablation, acoustic hemostasis, and propagation through porous bones will be illustrated. [Work supported by US Army Medical Research Acquisition Activity and NIH NCRR Grant R21 RR 21472.]

Session 2pBBb

Biomedical Ultrasound/Bioresponse to Vibration: Biomedical Imaging and Microscopy

Suk Wang Yoon, Cochair
Sung Kyun Kwan Univ., Dept. of Physics, 300 Chunchun dong, Jangan Ku, Suwon 440-746, Rep. of Korea

Yoshifumi Saijo, Cochair
Tohoku Univ., Dept. of Medical Engineering and Cardiology, 4-1 Soiryomachi, Aoba-ku, Sendai, 980-8575, Japan

Contributed Papers

3:30

2pBBb1. Temperature dependence of acoustic properties of human tissues with a high-resolution ultrasound backscatter microscope. Jee Woo Park (Acoust. Res. Lab., Dept. of Phys., SungKyunKwan Univ., Suwon 440-746, Republic of Korea), Jiehui Liu, Byoung Nam Kim, and Suk Wang Yoon (SungKyunKwan Univ., Suwon 440-746, Republic of Korea)

The B-mode *in vitro* images of human skin and ocular tissues are obtained using a high-resolution ultrasound backscatter microscope (UBM) at the frequencies of 40, 80, and 100 MHz. It shows that a high-resolution UBM can visualize more detailed superficial microstructures of skin and eye at higher frequencies. In order to improve high-frequency ultrasound images, we have investigated the acoustic properties, such as sound velocity, sound attenuation, and backscatter coefficient, of human tissues with this UBM system over the frequency range from 25 to 110 MHz. Their acoustic properties are also investigated as a function of temperature in the range from 20 °C to 65 °C, which may provide bases for ultrasound image monitoring of thermal therapy. [Work supported by Strategic National R&D Program of Ministry of Commerce, Industry and Energy (MOCIE), Republic of Korea.]

3:45

2pBBb2. Calibration for biological acoustic impedance microscope. Naohiro Hozumi (Aichi Inst. of Technol. 1247 Yakusa-cho, Toyota, 470-0392 Japan), Satoshi Terauchi, Ayumi Kimura, Masayuki Nagao, Sachiko Yoshida (Toyohashi Univ. of Technol.), Kazuto Kobayashi (Honda Electron.), and Yoshifumi Saijo (Tohoku Univ.)

The presentation deals with the calibration for a biological acoustic impedance microscope. A block of tissue was placed on a plastic substrate, and an acoustic pulse focused on the interface was transmitted from the rear side of the substrate. The reflection was received by the same transducer and interpreted into local acoustic impedance. An acoustic impedance image was obtained by mechanically scanning the transducer. A reference material was simultaneously observed in order to perform calibration. A water droplet was employed as a reference material, however, it often entered into the interface between the substrate and tissue, leading to the reduction of the image quality. Therefore, several kinds of compounds were investigated. Silicone rubber was relatively homogeneous and stable in acoustic impedance, after a certain period of curing time. However, its dispersion with the scale of focal spot size was as large

as 0.02 Ns/cm^3, suggesting that a sufficient number of points should be measured and averaged in order to obtain a precise reference signal. Finally, change in local acoustic impedance of cerebella tissue of a rat as a function of postnatal days was quantitatively assessed for each independent layer like molecular layer, granular layer, and white matter.

4:00

2pBBb3. Ultrasonic evaluation of blood coagulation from stroke patients. Ying Li, Dong-Guk Paeng (Oceanic Information and System Eng., Ocean College, Cheju Natl. Univ., JEJU, Korea, 690-756), Sitaramanjaneya Anjaneya Reddy Guntur, Min Joo Choi, Young Ree Kim, and Jae-Chul Choi (Cheju Natl. Univ., JEJU, Korea, 690-756)

It is already known that the ultrasound speed and attenuation coefficient are increasing during the blood clotting process. A pulse echo system was set up to measure the sound speed and attenuation coefficient of human blood at 37 C. From ten healthy volunteers, the blood was taken to measure sound speed and attenuation coefficient during blood coagulation without and with heparin. For ten stroke patients, sound speed and attenuation were measured with aPTT or PT for comparison. The variation of sound speed and attenuation coefficient of blood after adding heparin was small. Sound speed of plasma from healthy volunteers was increased to 7 from 1535 m/s over 50 min without heparin, but to 1 m/s during the first 5 min for heparin-added blood. Sound speed of plasma from stroke patients was changed little compared to that from healthy volunteers without heparin. Attenuation coefficient was increased linearly during plasma coagulation for both volunteers and stroke patients, but its temporal slopes were much smaller for stroke patients. The blood samples from the stroke patients that investigated how heparin affects sound speed and attenuation as a function of time before and after taking the heparin and the results will be discussed.

4:15

2pBBb4. Correlation between the degree of hemolysis and acoustic impedance, speed of sound, and attenuation in flowing whole blood. Adam White and Constantin-C. Coussios (Inst. of Biomed. Eng., Dept. of Eng. Sci., Univ. of Oxford, 17 Parks Rd., Oxford OX1 3PJ, UK, constantin.coussios@eng.ox.ac.uk)

The effect of increasing levels of mechanical hemolysis on the propagation of 15-MHz ultrasound in flowing whole blood at 37 °C was investigated. An experimental setup was developed, which enabled the continuous measurement of attenuation, speed of sound, and acoustic impedance

in porcine whole blood under conditions of laminar flow and mean shear rates well above the threshold for red cell aggregation. Starting with an initial hematocrit of 35%, an increasing volume fraction of the blood in the flow loop was hemolyzed by exposure to very high shear stresses. The attenuation coefficient at 15 MHz was found to vary between 1 dB/cm for 0% hemolysis (no red cell damage) to 5 dB/cm for 75% hemolysis, remaining approximately constant for higher hemolysis levels. Speed of sound decreases from 1561 m/s for healthy blood to 1545 m/s for 80% hemolysis. Finally, acoustic impedance was also found to vary from a minimum of 1575 Kg/m^2 to 1589 Kg/m^2 between 15% and 90% hemolysis. These results show great promise towards solving the inverse problem, namely developing a technique capable of deducing the degree of hemolysis noninvasively and in real time based on measured changes in ultrasound parameters.

4:30

2pBBb5. Multimodality tissue-type imaging of prostate cancer. Ernest Feleppa, Shreedevi Dasgupta, Sarayu Ramachandran (Riverside Res. Inst., 156 William St., New York, NY 10038, feleppa@rrinyc.org), Christopher Porter, David Dail, and Marc Lacrampe (Virginia Mason Medical Ctr., Seattle, WA 98101)

Unlike many tumors, prostate cancer cannot be imaged reliably by any single, existing, conventional modality. This makes guiding biopsies, targeting therapy, and monitoring effects problematic. Therefore, our objective is to improve significantly the depiction of prostate cancers using multimodality, multifeature techniques combined with nonlinear classifiers such as artificial neural networks (ANNs) and support vector machines (SVMs). Our previous studies show encouraging potential for tissue typing and imaging based on ultrasound spectrum analysis using ANNs and SVMs that produce ROC-curve areas of 0.84 to 0.88 compared to 0.64 for conventional ultrasound imaging. Similar studies by others have shown that magnetic-resonance spectroscopy can produce comparable ROC-curve areas based on the ratio of choline and creatine to citrate. Because these two modalities sense entirely distinct properties of tissue, i.e., microscopic mechanical properties by ultrasound and chemical properties by magnetic resonance, their combination should markedly improve classification performance in distinguishing between cancerous and noncancerous prostate tissue. We have undertaken preliminary studies incorporating these two modalities. These studies demonstrate the feasibility of coregistering features from each modality and generating new, combined-modality images that exploit the attributes of both modalities.

4:45

2pBBb6. Validating the theoretical framework relating ultrasonic spectrum-analysis parameters to scatterer properties. Shreedevi Dasgupta, Ernest Feleppa, Jonathan Mamou (Riverside Res. Inst. 156 William St., NY 10038), and Mark Rondeau (Cornell Univ., Ithaca, NY)

Existing theoretical frameworks for ultrasound spectrum analysis relate properties of backscattered echo signals from tissues to scatterer properties. However, the reliance of these frameworks on numerous key assumptions may not be appropriate for evaluating tissue properties. The objective of our studies is to validate those assumptions or reduce dependence on them so that estimates of scatterer properties can be made with greater confidence. Ultrasound rf data were acquired from isolated, uveal melanoma cells suspended in liquid media using center frequencies of 10 and 35 MHz. Average cell size was 12 μm and concentrations ranged from 1.3×10^8 cells/ml to 2.0×10^6 cells/ml. Theoretical spectral-parameter values predicted by the Lizzi framework were calculated using known values for size and concentration and published values for sound propagation velocity and acoustic impedance. Normalized spectral parameters were computed from the experimental rf data using standard Fourier methods. Predicted and experimental values for the slope, midband, and intercept parameters agreed to within 11%. Initial results tend to validate the theory at these scatterer concentrations. Future work will provide more insight into scattering behavior under a greater variety of conditions, e.g., scatterer sizes and concentrations.

5:00

2pBBb7. Alias-free estimation of blood velocity using wideband pulse transmission. Naohiko Tanaka (Faculty of Systems Eng., Shibaura Inst. of Technol., 307 Fukasaku, Saitana 337-8570, Japan, tanaka@se.shibaura-it.ac.jp)

Blood velocity information that is obtained from a pulsed Doppler system is applied to clinical diagnoses of the circulatory system. However, the conventional pulsed Doppler technique has a fundamental limitation on Doppler velocities: blood velocity is measured erroneously when it is greater than the Nyquist limit. This phenomenon is called aliasing. This study is aimed at expanding the range of measurable blood velocities while avoiding aliasing. A new technique proposed here is based on wideband pulse transmission and phase correction. A series of short pulses with a wide frequency band is transmitted in this technique. Received echoes are Fourier transformed and converted to the cross spectrum. The phase of the cross spectrum is proportional to the frequency and the blood velocity. When the blood velocity is greater than the Nyquist limit, a part of the phase of the cross spectrum is larger than $\pm \pi$ and is wrapped in $\pm \pi$. The wrapped phase data are corrected using wideband echo information to avoid aliasing. Then the gradient of the corrected phase is estimated and converted to the blood velocity. Simulation results show that the proposed technique triples the velocity range.

5:15

2pBBb8. Continuous delay estimation with polynomial splines. Gianmarco Pinton and Gregg Trahey (Duke Univ., 136 Hudson Hall, Box 90281, Durham, NC 27708, gfp@duke.edu)

Delay estimation is used in biomedical ultrasonics to determine blood flow, phase aberration corrections, and elastographic images. The accuracy of the algorithms developed to determine these delays depends on noise, bandwidth, and delay range. Subsample delay resolution is required in most cases. Two new algorithms that use cubic polynomial splines to continuously represent the delay and the apply pattern matching functions are introduced. The continuous analytic expressions for the pattern matching functions are used to calculate the delay. Normalized cross correlation and the sum of the squared error are implemented. These algorithms are compared to conventional delay estimators, such as discrete normalized cross correlation, autocorrelation, and to another spline-based method. Simulations compare the performance across noise, delay, and bandwidth. The proposed algorithms perform better, in terms of bias and jitter. For example, the jitter at a 0.5 subsample shift is 2.3×10^{-3} subsample shifts for interpolated normalized cross correlation, and 3.5×10^{-4} and 4.0×10^{-4} for the two proposed algorithms. The computational requirements of the new algorithms are also considered and are shown to be competitive with the interpolated normalized cross correlation at low interpolation values and smaller at large interpolation values. [This work was supported by NIH grants R01-HL075485 and 1R01-CA114093-02.]

5:30

2pBBb9. Erythrocyte aggregability from human, horse, and rat blood by ultrasound backscattering measurements. Kwon-Ho Nam (Interdisciplinary Postgraduate Program in Biomed. Eng., Cheju Natl. Univ., 66 Jejudaehakno, Jeju-si Jeju-do, South Korea, 690-756), Dong-Guk Paeng, Tae-Hoon Bok, and Min Joo Choi (Cheju Natl. Univ., 66 Jeju-si Jeju-do, South Korea, 690-756)

A pulse-echo system was used in a cylindrical chamber with a stirring magnet and a 5-MHz focused transducer for studying erythrocyte aggregation from human, horse, and rat blood. The highest backscattering power was measured from horse blood and followed by human and rat blood both at stasis and under shear forces. In human and horse blood, the backscattered power (BP) was decreased with stirring speed and saturated within 1 min. The BP was increased to 20 dB when blood flow was stopped. Since the aggregation of rat erythrocytes in autologous plasma is very low and barely detectable by the ultrasound BP, rat plasma was

replaced by phosphate buffered saline containing polyvinylpyrrolidone 360 in order to enhance the aggregation. In conclusion, this simple *in vitro* ultrasound measurement system requires a small amount of blood, which may provide an alternative method to investigate the mechanisms and

physiological roles of erythrocyte aggregation. Furthermore, the rat erythrocytes in polymer solution may be useful for an automatic measurement system to examine the bioactive materials that affect aggregation in preclinical studies.

Session 2pEA

Engineering Acoustics: Ultrasonic Functional Devices II

Gary W. Elko, Cochair

mh acoustics LLC, 26 Blackburn Pl., Summit, NJ 07901

Jun Kondoh, Cochair

Shizuoka Univ., Graduate School of Science and Technology, Johoku 3-5-1, Hamamatsu-shi, Shizuoka 432-8561, Japan

Invited Papers

1:00

2pEA1. Development of shear-horizontal–surface-acoustic wave (SH-SAW) sensor for liquids. Jun Kondoh (Grad. School of Sci. and Technol., Shizuoka Univ., Johoku 3-5-1, Hamamatsu-shi, Shizuoka 432-8561, Japan) and Showko Shiokawa (SAW&SPR-Tech. Corp., Hamamatsu-shi, Shizuoka 432-8011, Japan)

Acoustic-wave devices have been applied as chemical and physical sensors for gas and liquid phases. Since the surface acoustic wave (SAW) on a 36°-rotated Y-cut X-propagating LiTaO$_3$ single crystal (36YX.LT) is in the shear horizontal (SH) mode, SH-SAW devices have been applied as liquid-phase sensor. In this paper, we present fundamental properties of the SH-SAW sensor and its applications. When a liquid is loaded on the SH-SAW propagation surface, evanescent fields of the SH displacement and piezoelectric potential penetrate into the liquid. Evanescent fields are influenced from the liquid, so phase velocity and amplitude of the SH-SAW are perturbed. By detecting the change, liquid-phase sensor is realized. Based on the theoretical calculation, the SH-SAW sensor is influenced by liquid properties, i.e., density and viscosity products, relative permittivity, and conductivity. For simultaneous detection of these properties, a three-channel SH-SAW sensor is fabricated. Evaluation of mixed solutions of glucose and ethanol is carried out using the sensor. The results indicate that the three-channel SH-SAW sensor can be applied to a fermentation monitoring system.

1:20

2pEA2. Lead zirconate titanate thick-film ultrasonic transducer for 1- to 50-MHz frequency bands fabricated by hydrothermal method. Mutsuo Ishikawa, Minoru Kurosawa (Interdisciplinary Grad. School of Sci. and Eng. Tokyo Inst. of Technol. Kurosawa lab. G2-32, 4259 Nagatutamati, Midori-ku, Yokohama, Kanagawa 226-8502), Tomohito Hasegawa, Hiroshi Kitsunai, Akito Endoh, Shinichi Takeuchi (Toin Univ. of Yokohama), Takeyoshi Uchida, and Tsuneo Kikuchi (Natl. Inst. of Adv. Industrial Sci. and Technol.)

This prototype ultrasonic transducer had about a 30-μm thick hydrothermal PZT layer on a 50-μm thick titanium substrate. An active area of this transducer was 3.5 mm in diameter. Material constants of this hydrothermal PZT thick film were measured. Electromechanical coupling coefficient kt was over 0.25 and acoustic impedance was 6 Mrayl. A mechanical resonance frequency in the thickness mode of the prototype ultrasonic transducer was approximated by a mason's equivalent circuit using the thickness and the material constants of the hydrothermal PZT film and the Ti substrate. The prototype ultrasonic transducer had a double broad peak spectrum of about 15 and 30 MHz in simulation. Also, these broad peaks were a half wavelength and a quarter wavelength resonance frequency in a combined structure of the hydrothermal PZT thick film and the Ti substrate. The prototype ultrasonic transducer transmitted ultrasound in water. Experimental results were the prototype ultrasonic transducer radiated ultrasound at the frequency range from 1 to 50 MHz as wideband using respective resonance frequency. A reason for the wideband characteristics radiation of ultrasound was the prototype ultrasonic transducer has low acoustic impedance and double broad peak.

Contributed Papers

1:40

2pEA3. Equivalent network representation for length-extensional vibration modes in a side-plated piezoelectric bar with a varying parameter. Ken Yamada (Dept. of Electron. Eng., Faculty of Eng., Tohoku Gakuin Univ., 1-13-1 Chuo, Tagajo 985-8537, Japan)

Equivalent network representation is derived for length-extensional vibration modes in a side-plated piezoelectric bar with varying parameters along it. In contrast to findings of a former report [Yamada *et al.*, IEEE Trans. UFFC **48**(2), 613–616 (2001)], which dealt with the longitudinal

piezoelectric effect, the bar in this study is presumed to be excited through a transverse piezoelectric effect. The network comprises an electrical transmission line of length corresponding to the bar length and an electrical port with a capacitor that is linked to the line at both its ends via two ideal transformers. Frequency characteristics of the input electric admittance of the bar with a sinusoidally varying piezoelectric parameter, which shows the unique feature of the vibrator, is demonstrated using equivalent network analysis. The possibility of applying this network for modeling the interdigital transducer for excitation and detection of surface acoustic waves (SAWs) is discussed.

1:55

2pEA4. Measuring micro liquid volume by acoustic resonant technique. Makoto Wakabayashi, Daisuke Koyama, Kentaro Nakamura, and Sadayuki Ueha (P&I Lab. Tokyo Technol., 4259-R2-26, Midori-ku, Yokohama, Kanagawa 226-8503, Japan)

A method to measure micro liquid volume less than 300 ln with the accuracy of 0.1% is highly needed in biogenetic and biochemical industries. The authors propose a novel method to estimate micro liquid volume in a small well using acoustic resonance. In our experiments, liquid is in a container 6.5 mm in inner diameter and 9 mm in height, and the sound pressure near the well is measured by sweeping the illuminating sound field from 6 to 11 kHz. According to the relationship between the resonance frequency and the volume, we can successfully estimate the small volume with the accuracy of 0.1% and an excellent repeatability. For simultaneous multipoint measurements, we experimentally investigate the effect of adjacent wells on resonant frequency of the target well. It is revealed that the resonance frequency shift due to the mutual coupling between the adjacent wells is negligible if the well interval is less than 5 mm.

2:10

2pEA5. A multidimensional overview for ultrasonic motor driven loudspeaker development. Hirokazu Negishi, Juro Ohga (Dept. of Electron. Systems Eng., Essex Univ., 1-37-14 Koyabe Yokosuka 238-0026, Japan, hiro.note@jcom.home.ne.jp), and Juro Ohga (Shibaura Inst. of Technol., Tokyo 108-8548, Japan)

Twelve years of history for the development activity of the ultrasonic motor driven loudspeaker show a multidimensional aspect. Three main axes are geography, frequency, and mechanical impedance. Geography came first since initial conception was in the UK, yet the prototyping was in Japan. The origin was an idea for a stepping motor driven digital loudspeaker that was to address improving abnormalities coming from the fundamental resonant frequency, the so-called "F0." Consultation revealed the idea was weak but unexpectedly "ultrasonic motor" was recommended. This led the prototyping to be in Japan, where the invention and development of the motor took place. The second is frequency. For audible sound reproduction, inaudible ultrasonic frequency is employed. Although combination of these two regions in frequency is not new for sound reproduction, a couple between the bottom end of the audible sound reproduction and the ultrasonic motor is completely new for this arena. The third is mechanical impedance since convention is low whereas this one is high. Actually, this is the core of the attempt as most abnormalities around "F0" stem from weak mechanical impedance. Other axes also exhibit contrast against conventions such as linearity, controlling, driving, and lifetime of the motor itself, as well as configuration of the sounding mechanism.

2:25

2pEA6. Loudspeaker using continuous revolution of a piezoelectric ultrasonic motor. Yuta Ohnuma, Yusuke Iwaki, Juro Ohga (Shibaura Inst. of Technol., 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan, m106020@sic.shibaura-it.ac.jp), Hirokazu Negishi (Essex Univ., Yokosuka 238-0026, Japan), and Kazuaki Maeda (TOA Corp., Japan)

Volume displacement of a direct radiator loudspeaker diaphragm is inversely proportional to the square of its signal frequency. The lowest frequency of a flat response is restricted by the diaphragmfs fundamental resonance frequency. Therefore, a loudspeaker diaphragm for radiation of a low-frequency signal shall have large amplitude and low resonant frequency. The authors propose the application of a piezoelectric ultrasonic motor for this purpose. It produces a completely controlled large output force because its output mechanical impedance is much greater than any conventional transducer or motor. An ultrasonic motor whose stator is connected to a direct-radiator loudspeaker cone diaphragm by a rod and whose rotor is loaded by a heavy metal ring rotates with a constant veloc-

ity. Modulation of the velocity by the audio signal causes driving force to the diaphragm because a heavy ring with large inertia tends to rotate with constant velocity. This paper reports construction of a practical experimental model and its output sound pressure characteristics.

2:40–3:00   Break

3:00

2pEA7. Removal of thin hard coating on glass substrate by collision of nanometer-sized particles accelerated in high-intensity ultrasonic field. Naoya Kudo and Kazunari Adachi (Faculty of Eng., Yamagata Univ., 4-3-16 Jounan, Yonezawa, Yamagata, 992-8510 Japan, ere69123@dip.yz.yamagata-u.ac.jp)

The aim of this study is to establish a fundamental recycling technology for optical lenses. In industry, a common conventional technique for cleaning the surface of a small object such as an optical lens is through the use of ultrasonic cavitation. However, the intensity of shock waves caused by the cavitation is difficult to control. The authors have considered that the substances that adhere to lens surfaces, such as hard optical coatings and stains, are removable using collisions, against the lens surface, of many tiny particles that are accelerated by a high-intensity ultrasonic field. Some physical parameters that dominate the removal effect have been changed systematically to illustrate the surface-refreshing process for lenses in the experiments. The authors worked to experimentally clarify the particle collision effect on the removal. In addition, the acoustic field was analyzed using finite-element method. The authors have also attempted to elucidate the principle of the removal technique through comparison of experimental results with those obtained through numerical simulation. A precise elucidation, however, remains as an important problem to be solved.

3:15

2pEA8. Atomization of high-viscosity materials by one-point convergence of sound waves radiated from an aerial ultrasonic source using a square transverse vibrating plate. Hikaru Miura (Dept. of Elec. Eng., College of Sci. and Technol., Nihon Univ., 1-8-14, Kanda Surugadai, Chiyoda-ku, Tokyo 1018308, Japan)

It is difficult to atomize high-viscosity materials such as oil without contact; however, it is possible to do so using intense aerial ultrasonic waves. As a sound source for that purpose, an aerial ultrasonic source with 20-kHz resonance frequency was used along with a square transverse vibrating plate that has the nodal mode of vibration amplitude in the shape of a lattice. The source radiates four-directional sound waves in the far field. To obtain high sound pressure in the near field, radiated sound waves were converged using several flat boards and four parabolic reflective boards placed near the source. Sound pressure reaches 6.5 kPa (S.P.L. 170 dB), a very high level, at the convergent point. Scattering and atomization of the high-viscosity materials were examined at the point of convergence. Results show the following. Atomization by a sound wave is possible in the case of high-viscosity materials. The higher the sound pressure, the shorter the beginning time of atomization becomes. The distribution of the atomized particles diameter is 0.05–0.22 mm at the movement viscosity of 10 000 cSt.

3:30

2pEA9. Ultrasonic complex vibration converters using different sound velocity metal ring pairs with slanted thickness. Jiromaru Tsujino, Tetsugi Ueoka, Goh Kishimoto, and Rohei Karatsu (Kanagawa Univ., 3-27-1 Rokkakubashi, Kanagawa-ku, Yokohama 221-8686, Japan)

The ultrasonic complex vibration systems are effective for various high power applications. New complex vibration converters using different sound velocity metal ring pairs with slanted thickness are proposed and studied. Two different sound velocity rings with a slanted surface along the diameter are superimposed. The converter ring pair may be installed in any place of a longitudinal vibration system and driven by a longitudinal vibration source. The system is driven transversally according

to traveling time difference between both sides of the converter. The vibration rod vibrates in elliptical to circular locus in the case where these two natural frequencies are adjusted as phase difference is near to 90 deg. Thus, various one-dimensional complex vibration systems are obtained. Piezo-ceramics cut or polarized slanted may be used. The 18-mm-diam converter for 27-kHz system consists of a phosphor bronze and stainless steel rings. The rings are cut slanted, 1.5 to 6.5 mm in thickness at outer diameter, and superimposed to an 8-mm-thick ring. Two or three converters are available for more complicated vibration modes. Using 27- and 150-kHz complex vibration systems, 1.0-mm-thick aluminum plates and coated thin copper wires for IC cards were welded directly with material strength.

3:45

2pEA10. Nondestructive inspection system for corrosion detection in fuel tank liners. John W. Doane, Michael D. Gray, and Jayme J. Caspall (Georgia Inst. of Technol., 771 Ferst Dr., Atlanta, GA 30332, jd37@mail.gatech.edu)

An ultrasonic, nondestructive inspection system for the early detection of corrosion in fuel tank aluminum liners was developed and tested. A Labview-controlled scanner, which rode on the tank's resident rib braces, housed a 10-MHz transducer and a miniature video camera. The bidirectional (axial, azimuthal) scan resolution was controlled by adjusting the surface standoff distance of the focused transducer to produce variable-area interrogation sites. The phenomenon of increased backscatter from corrosion-roughened surfaces was exploited to detect the presence of corrosion in its early stages. The transducer acted as both transmitter and receiver, capturing the echoes from the subject surface while the video camera captured an image of the surface under acoustic inspection. The acoustic and visual data were overlayed to form a composite image of the entire inspection surface. The system was field-tested using two tanks pulled from active duty at the Marine Corps maintenance facility in Cherry Point, NC. [Work supported by JCAA.]

4:00

2pEA11. Evaluation of the static stress exerted on the clamping bolt in the structure of a bolt-clamped Langevin-type transducer. Toru Takahashi and Kazunari Adachi (Adachi Lab., 8-312, Faculty of Eng., Yamagata Univ., 4-3-16 Jonan, Yonezawa, Yamagata, 992-8510, Japan, g02239@dipfr.dip.yz.yamagata-u.ac.jp)

Bolt-clamped Langevin-type transducers (BLTs) must possess various characteristics depending on the field in which they are used. It has already been clarified by the authors that prestress or bearing stress imposed on piezoelectric elements in a transducer should be sufficiently large to compensate for their low tensile strength. For this study, we measured the strain on the clamping bolt surface using strain gauges pasted on it to evaluate the static prestress that changes with the turning angle of the metal block in clamping. We calculated the strain for some BLTs of identical shape using the finite-element method (FEM) and measured it to

confirm the validity of the prestress calculation that was achieved previously by the authors. Unfortunately, a large difference in the strain between the calculated and measured results was observed. We cited three possible reasons for that huge discrepancy. Among them, plastic deformation of the screw thread might most influence the result. As long as the design problem of BLTs pertaining to the prestress remains unresolved, we cannot take full advantage of the excellent electromechanical transduction efficiency of BLTs for high-power ultrasonic applications.

4:15

2pEA12. Multilayered high-performance transducer using polyurea thin film. Marie Nakazawa, Masaya Tabaru, Daisuke Koyama, Kentaro Nakamura, Sadayuki Ueha (Tokyo Tech., 4259 R2-25, Nagatsuta, Midori-ku, Yokohama, Kanagawa 226-8503, Japan, nakazawa@sonic.pi.titech.ac.jp), and Akihiro Maezawa (Konica Minolta M.G. INC., Tokyo, Japan)

Polyurea thin film exhibits useful characteristics as a high-frequency ultrasonic transducer such as a low acoustic impedance of 70% of that of PVDF, considerably high piezoelectric constants compatible with conventional piezoelectric polymers, and a higher Curie temperature. A polyurea thin film transducer working at over 100 MHz can be made on any surfaces through vapor deposition polymerization with a high controllability of thickness. To enhance the transducer performance, a multilayered configuration is studied in this study. A special twin-vacuum-chamber is used for laminating the polyurea layers and the aluminum electrodes alternately. We fabricate two- and three-layered transducers with 1.5-μm polyurea thin films as well as single-layer transducers. To evaluate the transducer performances, we measure the capacitances, the electrical admittances, the coupling factors, and the Q values. The pulse/echo measurements are also conducted to determine the transmission and receiving characteristics of the transducers using a glass plate as a reflector. The results of the pulse/echo experiments show that the received voltage amplitude of the multilayered transducer is higher than that of the single-layer transducer.

4:30

2pEA13. Analysis of fluid-borne coupling in transducer arrays. Kassiani Kotsidou and Charles Thompson (Ctr. for Adv. Computation and Telecommunications, Univ. of Massachusetts Lowell, 1 Univ. Ave, Lowell, MA 01854)

In this work, we will analyze fluid-borne coupling between adjacent elements of a transducer array. In the case of a transducer having low inertia mass, such coupling can result in a reduction in the radiation efficiency and the directional fidelity of the array. The coupling factor and the mutual radiation impedance have been investigated by Lord Rayleigh (1903), Wolff and Malter (1929), and R. I. Pritchard (1960) for the case of an array comprised of ideal piston radiators. These results will be used as a starting point for our work. The effect of the inertia impedance of the transducer and the control of the special distribution of the radiated field is considered.

Session 2pEDa

Education in Acoustics: Education in Acoustics for Children

Takayuki Arai, Cochair

Sophia Univ., Dept. of Electrical and Electronics Engineering, 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102, Japan

Ralph T. Muehleisen, Cochair

Illinois Inst. of Technology, Civil and Architectural Engineering, 3201 S. Dearborn St., Chicago, IL 60616

Chair's Introduction—1:00

Invited Papers

1:05

2pEDa1. Sound education in urban environment. Yoshio Tsuchida (Kanazawa Inst. of Technol., 7-1 Ohgigaoka, Nonoichi, Ishikawa, 921-8501, Japan)

There are various viewpoints for education and there are various levels for a sound. We have some cases to teach a ''sound.'' A certain person teaches a sound as music, another person teaches it as physics. But education is not teaching. It is more important to feel and enjoy. Then, one's sensitivity for a sound will be nourished. Experimental listening is called Sound Education based on soundscape. The case of a workshop on Sound Education is addressed in this paper. Sound Education finally aims at improvement of our environment. Many activities of Sound Education were devised by R. Murray Schafer [A Sound Education, 100 Exercises in Listening and Soundmaking, Arcana Editions]. Some activities were devised by Joseph Bharat Cornell as Nature Game [Sharing Nature With Children (Sharing Nature Series), Dawn Publications.] Sound Education has several aims. It differs according to the group that sponsors it. Here are the workshops executed by me. One aim for environmental preservation, and another one aim for community design.

1:25

2pEDa2. Music as a vehicle to do science in the elementary classroom. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809)

During the past several years, a number of workshops were conducted for public school teachers at the elementary levels. Training and materials were provided to help teachers use music as a vehicle to introduce science concepts. Tools used in these workshops include a long spring to introduce standing waves and the concept of harmonics, a monochord to relate frequency ratios to musical intervals, and simple computer software to introduce wave addition and spectral analysis. All of these will be demonstrated and ways of introducing them in the elementary classroom will be discussed. [These workshops were supported with ASA and Eisenhower funds.]

1:45

2pEDa3. Exciting demonstration in acoustics by high-school teachers' group: "Stray Cats." Kanako Ueno (Inst. of Industrial Sci., Univ. of Tokyo and Tech. Committee on Education in Acoust. of ASJ, Komaba 4-6-1, Meguroku, Tokyo, 153-8505, Japan), Takayuki Arai, Fumiaki Satoh, Akira Nishimura, and Koichi Yoshihisa (Tech. Committee on Education in Acoust. of ASJ, Meguroku, Tokyo, 153-8505, Japan)

In Japan, to get students interested in a subject, high-school teachers often form a group to share their ideas and inventions on education. ''Stray Cats'' is one of the most active groups in physics. The group has been proposing many exciting demonstrations, which were contrived to support students' learning process with intuitive understanding of physics. Here, instead of using commercial equipment, they developed simple teaching tools that show physical phenomena in an exciting and attractive way, using quite common materials and daily goods. For example, the velocity of sound is measured by a pipe filled with a gas (air, CO_2, helium, etc), setting a loudspeaker and two microphones in the pipe. Interference of sounds is demonstrated by two pipes with attached cone-shaped horns at one end, which collect a source sound in different phase, and merging them into one at the other end, which produce louder or quieter sound as a result of interference. Hitting or rubbing different length of aluminum rods aids students' understanding of longitudinal waves and transverse wave with the relationship between rod's length and pitch, as well as a pleasant experience with beautiful tones. These educational tools will be presented with videos taken with the Stray Cats group.

2pEDa4. An acoustics education outreach program for young children. Ralph T. Muehleisen (Civil and Architectural Eng., Illinois Inst. of Technol., Chicago, IL 60616, muehleisen@iit.edu)

With limited time for science in the preschool and primary classroom, and limited science education of many preschool and primary school educators, the need for science professionals to engage in outreach education is ever increasing. Additionally, there is an increasing need for children to be aware of the problems of noise exposure, especially with regard personal sound systems. In this talk, a one-hour acoustics education outreach program developed for preschool and early primary school children is presented. In the program, children are involved in a discussion of the human auditory system and the dangers of noise exposure, engage in listening exercises using a keyboard synthesizer, construct a small musical instrument, and finally play the musical instrument with the rest of the class in a marching band.

2pEDa5. Let children experience speech science. Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan)

An education in acoustics is important for students of different ages: college, high school, middle school, and potentially even elementary school students, for whom such hands-on models are the most important. Because speech communication is a basic human activity, even for children, we expect that even the youngest students to be interested in speech science. With this in mind, we were motivated to develop intuitive and effective models for teaching speech science to all types of students. Our physical models of the human vocal tract [T. Arai, J. Phonetic Soc. Jpn., **5**(2), 31–38 (2001)] have recently been installed in an exhibition in a museum in Japan. Children can watch, listen to, and touch the models. Our recent head-shaped and lung models [T. Arai, Acoust. Sci. Technol. **27**(2), 111–113 (2006)] are also useful for educating students, because they are intuitive to produce a systematic and comprehensive education from the lungs to the head. Furthermore, we recently designed a sliding three-tube (S3T) model that has a simple mechanism for producing several different vowels [T. Arai, Acoust. Sci. Technol (to be published)]. This model is being used for an activity at a science workshop for children. [Work partially supported by JSPS.KAKENHI (17500603).]

Contributed Papers

2pEDa6. Development of educational software that supports study of sound environment for children. Hiroyuki Sakai (Ctr. for Promotion of Excellence in Higher Education, Kyoto Univ., Yoshida-Nihonmatsu, Sakyo, Kyoto 606-8501, Japan), Tadashi Shirakawa (NPO Onbunken, Tokyo, Japan), and Munetaka Tanaka (Sound Process Design, Inc., Tokyo, Japan)

A sound map, on which observed sounds are drawn as visual symbols, is well known as one of the nature games. Sound educational software, cultivating children's ecological view through their daily experiences focusing on sounds, was developed by using the sound map concept. Children record observed sounds through a sound-walk in outdoors into PC, and learn the sound environment through their collaborative work including classification, analysis, and group discussions in a workshop or a class. In a traditional approach on the field of educational technology, a lot of teaching materials introducing novel ICTs (information and communication technologies) are developed, but it is rather important to support usefulness for users (children and teachers), promote motivations for their learning, and support the smooth learning progressive process. In this presentation, the development concept of the software for the sound environment study based on socio-cultural approach is discussed, while an actual use example of this software is demonstrated.

2pEDa7. Stimulating children's interest in acoustics using models and toys. M. G. Prasad (Dept. of Mech. Eng., Stevens Inst. of Technol., Hoboken, NJ 07030)

Sound is an important means for sensory perception. Children, being naturally curious and sensitive, respond very well to sound. Simple acoustical toys attract their curiosity and attention. The use of simple sound- or music-producing tools draws their attention and then they ask questions. In answering their questions simple principles of acoustics and scientific reasoning can be taught to them. Simple principles such as tone change due to change of volume in a resonator, the role of length in the sound produced by a tube, the role of damping in cutting down vibration and sound, etc. can be taught to the children. Also, the description of noise being unwanted sound can be demonstrated using simple mechanical models with comparison to desirable sound as music through instruments such as the flute. A model of the human ear can be used to describe the way the sound is perceived. The use of models and encouraging children to ask questions will further enhance their curiosity. The author will share his experiences gained through making presentations for elementary school children.

2pEDa8. Re-equipping the University of Sydney's acoustics laboratory for teaching after a fire. Ken Stewart and Densil Cabrera (Faculty of Architecture, Univ. of Sydney, NSW 2006, Australia)

For some decades, the University of Sydney has taught and conducted research in architectural and audio acoustics from the Faculty of Architecture's acoustics laboratory. In October 2005, a fire in the anechoic room destroyed the laboratory mainly through heat and smoke damage. Recovery has involved a reconsideration of the systems used for teaching and research, and this presentation outlines our approach to this. One consideration is that measurement systems should be relatively self-contained, so that setup time for measurements is minimized, and several setups can be used simultaneously for hands-on teaching. New equipment tends to be computer based, which has the advantages of much greater measurement power, automation, recording, and interfacing, but can be less intuitive for students to understand than individual analog components. This presentation compares the preexisting and renovated laboratory, especially in terms of teaching strategies in audio and acoustics using technical equipment. We present examples of previous and envisaged student projects to illustrate how such infrastructure can support learning.

2p WED. PM

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006

IAO NEEDLE/AKAKA FALLS ROOM, 3:45 TO 4:45 P.M.

Session 2pEDb

Education in Acoustics: Take 5's

Uwe J. Hansen, Cochair
64 Heritage Dr., Terre Haute, IN 47803

Kentaro Nakamura, Cochair
Tokyo Inst. of Technology, Precision and Intelligence Lab., 4259 Nagatsuta, Midori-ku, Yokohama, Kanagawa 226-8503, Japan

Chair's Introduction—3:45

Your favorite acoustics teaching devices. Bring a short demonstration, teaching device, or video, to share with your colleagues. A brief, descriptive handout is encouraged. Sign up at the door for a 5-minute slot. If you bring more than one demo, sign up for two non-consecutive slots.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006

MONARCH ROOM (ROYAL HAWAIIAN), 1:20 TO 4:15 P.M.

Session 2pMU

Musical Acoustics: Comparing Asian and Western Instruments II

Thomas D. Rossing, Cochair
Stanford Univ., CCRMA, Dept. of Music, Stanford, CA 94305

Tomoyasu Taguti, Cochair
Konan Univ., 8-9-1 Okamoto, Higashi-ku, Kobe, 658-8501, Japan

Invited Papers

1:20

2pMU1. Acoustical characteristics of Chinese stringed instruments and their Asian relatives. Akiko Odaka (Dept. of Musicology, Tokyo Natl. Univ. of Fine Arts and Music, 10-8 Uenokoen, Taito, Tokyo 110-8714, Japan) and Shigeru Yoshikawa (Kyushu Univ., Fukuoka 815-8540, Japan)

In Asia, there are many relative musical instruments that derive from the same origin or share a similar structure. In spite of their similarities, they show distinct differences from each other in terms of their musical usages, playing techniques, and acoustic effects. This varies with each musical genre, musical structure, social background, and acoustic taste swayed by those elements. People often add special devices to musical instrument or improve them so as to realize their acoustical ideas. For example, the shamisen and biwa, Japanese plucked lutes, have devices to allow strings to vibrate against the neck-wood, creating a reverberant high-frequency emphasis called sawari, which is of special value in each performance. A device that creates a similar effect, called jawari in Hindi, is also found in the sitar, a traditional stringed instrument of India. On the other hand, the Chinese sanxian, one of the relatives of the Japanese shamisen lacks this device. In this presentation, specific features of various musical instruments are introduced from both musicological and acoustical viewpoints. In particular, Chinese stringed instruments are focused on and compared with those of Japan, Korea, and India for investigation of their acoustical effects.

1:40

2pMU2. Vibro-acoustic measurement of several biwas. Tomoyasu Taguti (Konan Univ., 8-9-1 Okamoto, Higashinada, Kobe, Japan) and Kakuryo Tohnai (Nerima, Tokyo, Japan)

The biwa is a family of Japanese lutes that have been played for more than 1000 years. This paper reports some measurements on the vibro-acoustic characteristics of several biwas of different types. The main concern is the similarity and difference of the fukuju-to-table transfer mobilities among the individual biwas. Here, the fukuju is a kind of bridge whose shape and way of

installation differ distictively from those common to the majority of other lute families found in many countries in the past as well as in the present time. The measurement revealed that in most cases the fukujus have two major vibration modes, to be labeled the ''twist'' and ''flap'' modes, with the corresponding resonance frequencies in the regions of about 0.9 to 1.2 kHz and of 1.4 to 1.6 kHz, respectively, though the exact figures depend on the physical size and shape. Their resonance property takes an important roll to determine the fukuju-to-table transfer mobility as the coupled vibroacoustic system with the table. The fukuju-to-table mobilities were compared with the spectra of the corresponding plucked sounds.

2:00

2pMU3. The acoustics of the hammered dulcimer and similar instruments. David Peterson (Univ. of Central Arkansas, Conway, AR 72032)

The hammered dulcimer, an antecedent of the piano, apparently originated in the Middle East sometime around the 14th century. A typical instrument has 23 to 30 courses of strings, two to four strings per course, and is played with two wooden hammers. As in the yang qin and other Asian dulcimers, bridges separate some string courses into two notes. In American instruments the treble bridge creates a musical fifth, which facilitates the speedy playing of diatonic scales and also creates a strong harmonic structure to the tone. In response to player demand, modern builders have used improved or new materials (glues, woods, composites, plastic) and better design to create a lighter and less resonant instrument with a greater musical range. Dr. Peterson will demonstrate various playing styles.

2:20

2pMU4. Development of a carbon-fiber-reinforced composite for guitar soundboards: Investigation of bracing effect by experiment and simulation. Akira Okuda and Teruaki Ono (Faculty of Engineeing, Gifu Univ., Yanagido Gifu-City, Japan 501-1193, ono@cc.gifu-u.ac.jp)

A carbon-fiber-reinforced polyurethane foam composite whose physical and acoustic properties closely resemble those of wood for stringed-instrument soundboards was developed. A guitar with such a composite-made soundboard produced an almost identical tone to that produced with a wooden soundboard. The procurement of wood for soundboards, however, has not yet become very difficult. To accelerate the production and use of guitars with composite soundboards, it is necessary to create an attractive guitar that has characteristic and unique tones that cannot be realized on a wooden soundboard, making use of composite characteristics of design freedom. Wooden top boards, soundboards, have bracing that has traditionally and experientially been glued; however, the effects of bracing have never been clarified. In this study, the bracing effect on the frequency response and vibration mode in a guitar top board was investigated using experimentation and simulation. From the results obtained, a composite-made soundboard with no bracing, which produces more harmonics than a wooden one with bracing, was simulated as an example.

2:40

2pMU5. On the factors for detecting mistuned conditions and guiding their corrections for the Koto tuning. Tamaki Ando (Tokyo Natl. Univ. of Fine Arts and Music, 12-8 Ueno Park, Taitou-ku Tokyo, Japan 110-8714, an_tamaki@yahoo.co.jp)

By observations at the Koto tuning practice, it is realized that the bridge is moved in the period of about 400 ms, which is too short to feel the sensation of beat. Then, two kinds of listening tests were planned. The test sounds composed of two Koto sounds which should be tuned in one octave interval were presented at the same time like a part of actual Koto tuning process. The lower tone was at 196.9 Hz, the pitch of the higher one was varied in the range of 20 cents. Their length from the beginning was varied from 50 to 500 ms. The first experiment is for the judgment of the right or wrong tuning (higher or lower). The second one is for the tone color impression using 13 pairs of adjectives. The results of these experiments show that reasonable pitch judgments can be done for the test sounds longer than 200 ms by well-experienced Koto players, and the test sounds are clearly distinguished from each other by their tone color difference, whereupon in the Koto tuning by expert players, the direction and degree of mistuning are found to be judged at the initial stage of Koto sounds by the tone color impression.

3:00−3:15   Break

Contributed Papers

3:15

2pMU6. Acoustics of the *haegeum*, a Korean string instrument. Junehee Yoo (Dept. of Phys. Education, Seoul Natl. Univ., Seoul 151-742, Republic of Korea, yoo@snu.ac.kr) and Thomas D. Rossing (Stanford Univ., Stanford, CA 94305)

The Korean *haegeum*, closely related to the Chinese *erhu* and Japanese *kokyo*, dates back to the Goryeo Dynasty. It is played with a horsehair bow that is inserted between two strings attached to a small barrel-shaped resonator. A bow made up of a bamboo stick with lengths of horsehair is held by the right hand of the performer. The two strings are generally tuned to A3♭ and E4♭ and the resonator is tuned to about 300 Hz. The resonator has a paulownia wood plate at one end on which the bridge rests. By means of holographic interferometry, we compare modes of vibration in resonators of bamboo, ailanthus, birch, and ash.

3:30

2pMU7. Inharmonic sounds of bowed strings in Western music and Beijing opera. Chen-Gia Tsai (Grad. Inst. of Musicology, Natl. Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan, tsaichengia@ntu.edu.tw)

Due to the nonlinear bow-string interaction, inharmonic sounds can be produced by bowed strings. Such sounds are characterized by subharmonics, sidebands, and noises. Bow location, bow speed, and bow pressure on the string have been found critical to producing inharmonic sounds on the violin or the cello. In the accompaniment music of Beijing opera singing,

inharmonic sounds produced by the major bowed-string instrument (the jinghu) are sometimes called ''blossoming tones.'' They are played with a special touch technique of the left hand, and a chaotic behavior of string vibration arises from a nonlinear boundary condition with unilateral con-straints. The spectra of blossoming tones can be viewed as high-pass noises, with the cutoff frequency of about 1.6 kHz. Blossoming tones are sophisticatedly used for emotional expressions in Beijing opera. Their harsh quality recalls some pathological human voices.

3:45−4:15
Mini-concert of Koto music by Tamaki Ando

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006　　　　　　WAIANAE ROOM, 1:00 TO 3:45 P.M.

Session 2pPA

Physical Acoustics and Noise: Sonic Boom From Supersonic Aircraft and Projectiles II

Victor W. Sparrow, Cochair

Pennsylvania State Univ., Graduate Program in Acoustics, 316B Leonhard Bldg., University Park, PA 16802

Kazuyoshi Takayama, Cochair

Tohoku Univ., Biomedical Engineering Research Organization, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

Invited Papers

1:00

2pPA1. Realism assessment of sonic boom simulators. Brenda M. Sullivan (NASA Langley Res. Ctr., M.S. 463, Hampton, VA 23681, brenda.m.sullivan@nasa.gov), Patricia Davies (Purdue Univ., West Lafayette, IN 47907-2031), Kathleen K. Hodgdon (Penn State Univ., State College, PA 16804-0030), Joseph Salamone (Gulfstream Aerosp. Corp., Savannah, GA 31407), and Anthony Pilon (Lockheed Martin Aeronautics Co., Palmdale, CA 93599)

The perceived realism of booms reproduced in three simulators, at NASA, Lockheed Martin, and Gulfstream Aerospace, was studied. The NASA and Lockheed facilities are loudspeaker-driven, airtight concrete structures that enable reproduction of the entire spectrum, including the very lowest frequencies. The Gulfstream facility is enclosed within a trailer and consists of a folded horn as well as loudspeakers. This creates the sonic boom waveform as a traveling wave but is unable to reproduce the lowest frequencies. In the tests, subjects experienced in listening critically to real sonic booms heard reproductions in the simulators of sonic booms recorded outdoors. Tests comparing the NASA/Gulfstream simulators and the Lockheed/Gulfstream simulators used the same outdoor sonic boom waveforms, though each simulator uses its own preprocessing to equalize the waveforms for presentation. The tests showed that the presence of ground reflections, very low frequencies, or recording hiss were not factors in realism ratings. However, the inclusion of some of the postboom noise heard after the boom itself was found to be extremely important for a realistic boom simulation.

1:20

2pPA2. Sonic boom perception: Live versus simulator playback. Kathleen Hodgdon, Tom Gabrielson (Appl. Res. Lab., Penn State, University Park, PA 16802), Anthony Atchley (Penn State University Park, PA), Brenda Sullivan (NASA Langley Res. Ctr., Hampton, VA 23681), Joe Salamone, and Chris Babcock (Gulfstream Aerosp. Corp., Savannah, GA 31407)

A research team conducted a study at NASA Dryden Flight Research Center to compare the perception of real versus simulated booms. The field test included an assessment of noise measurement systems and a subjective jury trial on perception of sonic booms heard outdoors. Seventeen sonic booms were recorded under the flight path of four separate flights of three F-18s, using various microphones and recorders at multiple ground and ear level locations. Waveform analysis was conducted comparing the recordings as a function of location and measurement system. Eighteen subjects listened to live sonic booms, then rated the realism of the playback of a selection of those booms reproduced in the Gulfstream Supersonic Acoustic Signature Simulator II several hours later. The subjects also provided adjective descriptors of boom perceptions. The booms selected for the jury trial presented a variety of boom characteristics that included postboom noise. The subjects rated the playback as realistic, contributing to the validation of the simulator for use in future testing. Waveform comparisons and results from the field measurements and jury trial will be presented. [Work sponsored by FAA/NASA/TC PARTNER Center of Excellence.]

1:40

2pPA3. Preliminary work on the development of a lexicon for supersonic aircraft sound. Andrew Marshall, Patricia Davies (Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031, aimarsha@purdue.edu), Brenda Sullivan (NASA Langley Res. Ctr., Hampton, VA 23681), and Anthony Pilon (Lockheed Martin Aeronautics Co., Palmdale, CA 93599)

A series of tests was conducted in three sonic boom simulators to test how expert subjects perceived the realism of the sound playback. Subjects heard sounds (recorded sonic booms) in two simulators and rated them on a realism scale. On completion of this realism test, subjects were asked to write down descriptions of the sounds and also of real booms (indoor and outdoor) that they had heard in the past. They then listened to 14 sounds and, after each sound, wrote down adjectives and phrases describing the sound

characteristics. The 14 sounds were also analyzed and sound attribute metrics calculated for the main part of the boom as well as for the after-boom noise. The results of this analysis were compared to the words used to describe the sounds, and a preliminary set of words is proposed for inclusion in a lexicon for supersonic aircraft sound.

2:00

2pPA4. Human response to low-intensity sonic booms heard indoors and outdoors. Brenda M. Sullivan, Jacob Klos, Ralph D. Buehrle, David A. McCurdy (Structural Acoust. Branch, NASA Langley Res. Ctr., Hampton, VA 23681, brenda.m.sullivan@nasa.gov), and Edward A. Haering, Jr. (NASA Dryden Flight Res. Ctr., Edwards, CA 93523-0273)

A house on Edwards Air Force Base, CA, was exposed to low-intensity N-wave sonic booms during a 3-week test period in June 2006. The house was instrumented to measure the booms both inside and out. F-18 aircraft were flown to achieve a variety of boom overpressures from approximately 0.01 to 0.06 psf. During 4 test days, 77 test subjects heard the booms while seated inside and outside the house. Using the Magnitude Estimation methodology and artificial reference sounds, the subjects rated the annoyance of the booms. Since the same subjects heard similar booms both inside and outside the house, comparative ratings of indoor and outdoor annoyance were obtained. Preliminary results from this test will be presented.

2:20

2pPA5. Building structural acoustic response to aircraft sonic booms. Ralph D. Buehrle, Jacob Klos, Brenda M. Sullivan (Structural Acoust. Branch, NASA Langley Res. Ctr., Hampton, VA 23681), and Edward A. Haering, Jr. (NASA Dryden Flight Res. Ctr.)

As part of the NASA Low Boom/No Boom flight test project, a series of low-amplitude sonic-boom tests was completed over a 3-week period in June of 2006. This series of flight tests was designed to evaluate indoor/outdoor human subjective response, structural acoustic building response, and the effects of atmospheric turbulence for low-amplitude sonic booms characterized by overpressures in the nominal range of 0.1 to 0.6 pounds per square foot (psf). Low-amplitude sonic booms were generated by F-18 aircraft using dive trajectories to produce a range of overpressures at the Edwards Air Force Base housing area. In addition, straight and level supersonic flights were used to generate normal level (nominally 1.4 psf) sonic-boom overpressures at the housing area. This report will describe the structural acoustic building response measurements obtained during this flight test project. A single-family ranch-style home was instrumented with nearly 300 microphone and accelerometer sensors to determine the incident pressure and the building's structural acoustic response. Data were acquired for over 100 sonic-boom events. The test setup will be described and analysis of the measurements will be presented.

2:40

2pPA6. Modifications to the Gulfstream sonic boom simulator for the enhancement of low-boom simulations. Chris Babcock and Joe Salamone (Gulfstream Aerosp. Corp., P.O. Box 2206, Savannah, GA 31402)

A procedure is presented to enhance the simulation of measured low booms. Qualitative results show the SASSII, as originally designed, successfully reproduced the realism of full scale N waves. The simulator did not meet the realism level of the full-scale N waves when trying to reproduce the realism of lower amplitude sonic booms referred to as low booms. Data acquisition hardware was upgraded. Time alignment of each driver was performed with a different technique. Equalizer and amplifier programming was modified for each driver to provide a higher signal-to-noise ratio necessary to increase the realism of measured low booms in the simulator. A multipoint equalization was performed with a modified listening area. An iterative method rather than direct method was used to solve for the system of equations generated by the Weiner filtering scheme. The results are evaluated quantitatively by comparing the perceived loudness, A-weighted, and C-weighted sound-exposure levels of the full-fidelity waveform compared with the system output in the listening area.

3:00-3:15   Break

Contributed Papers

3:15

2pPA7. Prediction and measurement of sonic booms generated by the launch of a Minuteman II missile. Steven D. Pettyjohn (The Acoustcs & Vib.. Group, 5700 Broadway, Sacramento, CA 95820-1852)

The Air Force wanted to launch a Minute II missile full range. Predictions were 15.9 psf of over pressure at the sonic boom focus using similitude based on Saturn 5 data. Since this is at the threshold where single strength glass could break, a decision was made to predict over pressure using a NASA program. This program predicted a level of 16.1 psf over pressure, just 1.3% greater than the original estimate. Confirmation was made by measuring the sonic boom from the launch out of Vandenberg Air Force Base. Seven small boats equipped with sound-activated tape recorders were placed along the flight path with some boats placed at the predicted focus point based on the weather forecast provided by the Air Force. Based on work by A.D. Pierce, two distinct N-wave types were expected. One would represent the path from where the missile first reached sonic speeds and would propagate along a fairly shallow path with

respect to the earth's surface. The second wave type would arrive from a much higher elevation in the flight path and would propagate through many atmospheric layers. Weather forecasts were incorrect, but the results show the two shock wave types.

3:30

2pPA8. Modeling and measurements of nonlinear spherically divergent N waves in air. Vera A. Khokhlova, Mikhail V. Averiyanov, Petr V. Yuldashev (Dept. of Acoust., Faculty of Phys., Moscow State Univ., Moscow 119992, Russia, vera@acs366.phys.msu.ru), Sebastien Ollivier, and Philippe Blanc-Benon (LMFA, Ecole Centrale de Lyon, 69134 Ecully Cedex, France)

Propagation of nonlinear N pulses in relaxing medium was studied experimentally and numerically. An electric spark source was used in laboratory-scale experiments in ambient air to obtain spherically divergent shock pulses of 30-s duration and 1000-Pa peak pressure at 15-cm distance from the source. Pressure waveforms were measured by the broad-

band $\frac{1}{8}$-in. microphone at the distances varying from 15 cm to 2 m from the spark source. A numerical solution was obtained for an initial analytic N waveform at the same distances based on the modified Burger's equation for nonlinear spherical waves in dissipative medium with relaxation. The duration of the initial N wave was defined by matching the zero values of the corresponding analytic measured spectra; the amplitude was obtained by matching the experimental and theoretical nonlinear propaga- tion curves for the N-wave duration. The results of modeling were found to be in good agreement with the experimental data. A frequency response of the microphone was obtained at various distances as a ratio between the measured and modeled spectra and showed a good reproducibility. The relative effects of relaxation, nonlinearity, and microphone filtering on the N-waveform parameters are discussed. [Work is supported by RFBR and INTAS grants.]

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006

MOLOKAI ROOM, 1:00 TO 5:30 P.M.

Session 2pPP

Psychological and Physiological Acoustics: Physiology, Psychoacoustics, Noise, Prostheses (Poster Session)

Deborah A. Fantini, Cochair
Univ. of Essex, Dept. of Psychology, Wivenhoe Park, Colchester CO4 3SQ, U.K.

Kaoru Ashihara, Cochair
AIST, Tsukuba Central, 6 Tsukuba, 1-1-1 Higashi Ibaraki 305-8566, Japan

Contributed Papers

All posters will be on display from 1:00 p.m. to 5:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:30 p.m.

2pPP1. Umbo vibration in response to two-tone stimuli measured in human subjects. Diana Turcanu, Ernst Dalhoff, and Anthony W. Gummer (Univ. Tuebingen, HNO-Klinik, Elfriede-Aulhorn-Str. 5, 72076 Tuebingen, Germany)

The aim of these experiments was to noninvasively investigate cochlear mechanics near auditory threshold in human subjects by measuring the vibration response of the umbo. The vibration measurement setup consists of a custom-built laser Doppler vibrometer (LDV), employing a conventional heterodyne interferometer, for which all optical and electronic parts were tailored to the problem of measuring picometer-sized vibrations in the presence of large, extraneous movements in the order of 0.1 mm (e.g., heart beat, breathing, or swallowing). The cubic distortion product (DP) of the otoacoustic emissions (OAEs) in response to sound stimulation with two tones was measured as vibration on the umbo. For stimulus frequencies, f1 and f2, with f2/f1 = 1.2 and f2 = 4–9.5 kHz, and intensities L1 and L2, with L1 = 0.4L2 + 39 dB and L2 = 20–65 dB SPL, the vibration amplitudes were 1–8 pm in subjects ($N = 20$) with hearing loss no more than 25 dB SPL. Application of this vibration technique to a recently proposed method [P. Boege and T. Janssen, J. Acoust. Soc. Am. **111**, 1810–1818 (2002)] of objectively estimating hearing threshold by means of extrapolating DPOAE growth functions showed a standard deviation of the estimate of only 8.6 dB.

2pPP2. Optical imaging of neural activity to vocalized sounds in the guinea-pig auditory cortex. Junsei Horikawa, Takashi Hatta, Shunji Sugimoto (Toyohashi Univ. of Technol., Toyohashi 441-8580, Japan), Yutaka Hosokawa (Ryukyu Univ., Okinawa 903-0215, Japan), and Michinori Kubota (Tokyo Med. & Dental Univ., Tokyo 101-0062, Japan)

Spatiotemporal neural activity in response to vocalized sounds (calls) was investigated in the auditory cortex of anesthetized guinea pigs using an optical imaging technique with a voltage-sensitive dye (RH795). Guinea-pig calls were recorded digitally and presented to the ear from a loudspeaker located contralaterally to the recording cortex. Pure tone (PT), click (CLK), and white noise (WN) stimuli were used for comparison. The response in the primary (AI) and dorsocaudal (DC) fields of the auditory cortex to a PT appeared at the stimulus onset and in the corresponding frequency band, followed by nonsimultaneous lateral inhibition in the adjacent frequency bands. The responses to CLK and WN stimuli appeared in a wide area of the fields corresponding to the wide frequency components of the stimuli. The response to calls changed spatially and temporally because the components' frequencies changed over time. Strong responses occurred in the frequency bands corresponding to prominent instantaneous frequency components. These results suggest that activity to calls in AI and DC fields is formed by spatiotemporal excitatory and inhibitory interactions between the frequency bands responding to instantaneous frequency components in the calls. [Work supported in part by the 21st century COE program ''Intelligent Human Sensing'' from the MESSC, Japan.]

2pPP3. Extraction of sounds' time-frequency features in the auditory cortex. Shunji Sugimoto, Yuuta Suzuki, Hiroyuki Tanaka (Toyohashi Univ. of Technol., Toyohashi 441-8580, Japan, sugimoto@tutkie.tut.ac.jp), Michinori Kubota (Tokyo Med. & Dent. Univ., Tokyo 101-0062, Japan), and Junsei Horikawa (Toyohashi Univ. of Technol., Toyohashi 441-8580, Japan)

Although ''time-frequency feature extraction'' is a fundamental cue that is used to recognize speech sounds, its physiological mechanisms remain largely unknown. This research is aimed at understanding what time-frequency features of sounds are extracted by auditory cortical neurons. We investigated spatiotemporal response patterns in the guinea-pig auditory cortex, using an optical recording method (MiCAM Ultima; Brain Vision) with a voltage-sensitive dye (RH795). Experiments were performed under anesthesia (ketamine, 100 mg/kg, and xylazine, 25 mg/ kg). Pure tones (PTs), frequency-modulated (FM) sounds with various modulation rates, and their combinations were presented to the animals. Cortical activation was induced at the onset and offset of PTs along the

3122   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3122**

isofrequency contours. The onset response was followed by strong inhibition. In contrast, FM sounds evoked an additional activation moving across the isofrequency contours in the cortex, with locations corresponding to the instantaneous frequency input of the FM sweep. Such a response became more evident when the FM sweep rate increased. Based on these results, we discuss physiological mechanisms by which sounds with a rapid frequency modulation lead to "time-frequency feature extraction" in cortical neurons. [Work supported in part by the 21st century COE program "Intelligent Human Sensing" from the MESSC, Japan.]

2pPP4. An attempt to measure the hearing thresholds for high-frequency sounds by using the auditory brain-stem response.

Kazumasa Hosokawa (Musashi Inst. of Technol., 1-28-1 Tamatsutsumi, Setagaya-ku, Tokyo, 105-8557 Japan), Kaoru Ashihara (AIST, Tsukuba, Ibaraki, 305-8566 Japan), and Shogo Kiryu (Musashi Inst. of Technol., Setagaya-ku, Tokyo, 105-8557 Japan)

Recently, devices using high-power ultrasounds are increasing. However, influences of high level ultrasounds on human are not fully understood. Psychophysical studies show that the hearing thresholds exceed 80 dB SPL at a frequency above 20 kHz. It is reported that high frequency sounds above 20 kHz affect activities in the human nervous system. In order to examine whether the peripheral auditory system can be activated by high-frequency sounds, an attempt is made to measure the auditory brainstem response (ABR) for the frequency range between 4 and 20 kHz. In preliminary experiments, the ABR was measured for a tone pip with the center frequency of 14 kHz. However, the ABR was not clearly observed above 16 kHz. The thresholds determined by the ABR were approximately 40 dB larger than the psychophysical thresholds. The ABR was evoked by low-frequency components of the tone pip at the high presentation levels.

2pPP5. Distortion product otoacoustic emission fine structure as an early hearing loss predictor.

Karen Reuter and Dorte Hammershøi (Dept. of Acoust., Aalborg Univ., Fredrik Bajers Vej 7, B5, 9220 Aalborg Øst, Denmark, kr@acoustics.aau.dk)

Otoacoustic emissions (OAEs) are a promising method to monitor early noise-induced hearing losses. When distortion product otoacoustic emissions (DPOAEs) are obtained with a high-frequency resolution, a ripple structure across frequency can be seen, called DPOAE fine structure. In this study DPOAE fine structures are obtained from 74 normal-hearing humans using primary levels of L1/L2 = 65/45 dB. The subjects belong to groups with different ages and exposure histories. A classification algorithm is developed, which quantifies the fine structure by the parameter's ripple place, ripple width, ripple height, and ripple prevalence. Temporary changes of the DPOAE fine structure are analyzed by measuring DPOAE both before and after exposing some of the subjects to an intense sound. The characteristic patterns of fine structure can be found in the DPOAE of all subjects, though they are individual and vary from subject to subject within groups. The results do not indicate that the DPOAE fine structure alters with the state of hearing, as it is suggested in the literature. The data analysis is still in process at this stage.

2pPP6. The behavior of the distortion product otoacoustic emissions in response to contralateral stimulation by stationary and nonstationary signals.

Olimpia Szczepaniak (Div. of Rm. Acoust. and Psychoacoustics, Adam Mickiewicz Univ., Umultowska 85, 61-614 Poznan, Poland, oli0809@interia.pl)

The main aim of the study was to determine the DPOAE of the cochlea in the case of contralateral stimulation by stationary and nonstationary signals. The analysis was performed for $3f_1 - 2f_2$, $2f_1 - f_2$, $2f_2 - f_1$ intermodulation distortion products on a group of 20 normal hearing subjects (NH) and ten patients with low UCL (uncomfortable loudness level). The optimized primary tone level setting ($L_1 = 0.4L_2 + 39$ dB) was used. Each NH subject was characterized by the I/O function of otoacoustic emission (40–90 dB SPL) and dp-grams for the three levels of two-tone (85/75, 75/65, and 65/55 dB SPL) for frequencies between 1 and 12 kHz. Parameters for subjects with UCL were individually selected. The suppression of otoacoustic distortion products and the intergroup correlation of the slopes of I/O function with various contralateral stimulators were discussed. Results obtained from the experiment were compared between the two groups studied.

2pPP7. A functional MRI study on auditory feedback control of pitch.

Akira Toyomura (RIES, Hokkaido Univ., Sapporo 060-0812 Japan /JST, toyomu@es.hokudai.ac.jp), Sachiko Koyama, Tamaki Miyamoto, Atsushi Terao (Hokkaido Univ., Sapporo 060-0812, Japan), Takashi Omori (Tamagawa Univ., Tamagawa 194-8610, Japan), Harumitsu Murohashi, and Shinya Kuriki (Hokkaido Univ., Sapporo 060-0812, Japan)

Perception of one's own speech plays an important role in fluent speech production. In this study, we conducted a functional magnetic resonance imaging (fMRI) experiment to delineate the neural mechanism for auditory feedback control of pitch, using a transformed auditory feedback (TAF) method. Subjects ($n = 12$, right handed) were required to vocalize /a/ for 5 s while hearing their voice through a headphone. In the TAF condition, the feedback voice pitch was shifted randomly up or down from the original pitch two or three times. The subjects were required to hold the pitch of the feedback voice constant by changing the pitch of the original voice. In the non-TAF condition, the pitch of feedback voice was not modulated and the subjects merely vocalized /a/ continuously. A contrast between TAF and non-TAF conditions showed some significantly greater activations in the TAF condition ($p < 0.05$ corrected). In the right hemisphere, the supramarginal gyrus, the prefrontal area, the anterior insula, and the superior temporal gyrus showed greater activation. In the left hemisphere, only the premotor area showed greater activation. This result suggests that auditory feedback of pitch is controlled mainly by the right hemispheric network. [Work supported by JST.]

2pPP8. Mechanisms of bone-conducted ultrasonic perception assessed by electrophysiological measurements in human.

Seiji Nakagawa (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan, s-nakagawa@aist.go.jp) and Aya Nakagawa (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST)/ Suita Municipal Hospital, Osaka, Japan)

Several studies have reported that bone-conducted ultrasound (BCU) is perceived even by the profoundly sensorineural deaf; however, the mechanisms involved remain unclear. We previously reported some unique characteristics of BCU perception: (1) the pitch of BCU is about ten-odd kHz and is independent of its frequency, (2) the dynamic range of BCU is narrow, less than 20 dB, and (3) BCU mask 10–15-kHz air-conducted sounds. In this study, cortical magnetic field (N1m), middle latency (MLRs) and auditory brainstem responses (ABRs) and electrocochleogram (EcochG) evoked by BCU in human were recorded to clarify the neural pathway for BCU. Substantial MLRs and ABRs (wave-I–V) were evoked by BCU as well as by air-conducted sound. Although a clear N1m response was elicited by BCU and its equivalent current dipoles (ECDs) were estimated in the auditory cortices, ECD locations for BCU did not follow the tonotopic organization at the cortical level. Both wave I of ABRs and AP of EcochG were observed clearly, indicating that the auditory nerve was substantially activated. These results suggest that BCU goes through the normal auditory pathway—there are no special organs for BCU perception; however, unique processes may exist in the cochlea.

2pPP9. Interaction between the tonotopic and periodic information in recognition of "melodic" contours. Toshie Matsui, Chihiro Takeshima (Kyoto City Univ. of Arts, Grad. School of Music, 13-6, Kutsukake-cho, Oe, Nishikyo-ku, Kyoto 610-1197, Japan, m04905@kcua.ac.jp), and Minoru Tsuzaki (Kyoto City Univ. of Arts, Kyoto 610-1197, Japan)

It is known that the tonotopic and periodic aspects of sounds are separated in the auditory sense. To investigate the manner of representing these two aspects, a psychological experiment was designed. The stimuli are sequences of vowels, each of which changes its fundamental frequency and "size," i.e., the centroid of its formants, using a STRAIGHT vocoder. Participants are asked to respond if the transient patterns of the stimuli are "clockwise" or "counterclockwise" in a two-dimensional space, one of whose coordinates is the periodicity and the other of which is the size. Results indicate that the judgment was not so easy as in the case in which one would judge the transition patterns of points in a two-dimensional visual space. This inference implies that the two attributes might be represented as values on two completely independent scales, but not like the values in two coordinates of a single point.

2pPP10. A preliminary study of construction of a musical mood scale. Junko Matsumoto (Nagano College of Nursing, 1694, Akaho, Komagane, Nagano, 399-4117, Japan, matsumoto@nagano-nurs.ac.jp)

In this study to construct a mood scale that was specific to listening to music, appropriate items were selected preliminarily. Three hundred three university students were presented 70 adjectives by a questionnaire and asked to rate how frequent they feel each mood after listening to music in their daily lives on a four-point scale. A factor analysis were completed on the mood items and yielded five factors indicating suffering, comfortableness, solemnness and aggression, calmness, and sadness, respectively. Items with high loading on these factors were selected to construct a musical mood scale and any of them were modified because they might have an obscure meaning.

2pPP11. Sensitivity analysis to evaluate the benefits and risks of using prior information in adaptive psychometric procedures. Jeremiah Remus and Leslie Collins (Dept. of Elec. and Comput. Eng., Duke Univ., Box 90291, Durham, NC 27708)

Numerous adaptive psychometric procedures have been developed based on a Bayesian framework. In these procedures, a key component for estimating the targeted parameter and determining the next stimulus value is the parameter probability distribution. The presence of a parameter probability distribution introduces the opportunity to include prior information via initialization of the parameter probabilities. However, differences in the techniques for parameter estimation and stimulus selection may determine how various psychometric procedures benefit from accurate prior information or recover from inaccurate prior information. In this study, several adaptive psychometric procedures using a Bayesian framework were investigated through simulation of a psychophysical task. The simulations included prior information with varying degrees of accuracy and varying levels of initial confidence. Results suggest that the benefits of prior information are dependent on the specific psychometric procedure and the inclusion of prior information may influence the final choice of an appropriate psychometric procedure.

2pPP12. Observed variability in d' for yes-no and two-interval forced choice procedures. Walt Jesteadt, Lori Leibold, Samar Khaddam, Jessica Messersmith, and Melissa Krivohlavek (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

The sampling distribution of d' estimates can be determined from knowledge of the true d' and the number of trials per estimate [J. Miller, Percept. Psychophys. **58**, 65–72 (1996)]. Five subjects were tested in a fixed-level intensity discrimination task to obtain estimates of the observed variance of d' for both the yes-no and two-interval forced-choice (2IFC) procedures over a wide range of d' values. The stimuli were 2-kHz sinusoids, 300 ms in duration with 10-ms ramps, with differences in intensity ranging from 0.1 to 4.9 dB, centered on 70 dB SPL. An estimate of d' was computed for each of ten 100-trial blocks at each intensity difference for each procedure. Psychometric functions, based on 14 000 trials per function, were well fitted using intensity differences expressed in units of ΔL and were steeper for 2IFC than for yes-no. The variance of d' estimates exceeded the variance of the sampling distribution by a factor of 2.1 for yes-no and a factor of 3.1 for 2IFC, with learning effects contributing to the additional variance in some subjects. For a given value of d', the variance of d' estimates was approximately equal in yes-no and 2IFC. [Work supported by NIH/NIDCD.]

2pPP13. Neural correlates of psychometric functions for auditory detection. Huanping Dai (Div. of Commun. Disord., College of Health Sci., Univ. of Wyoming, Laramie, WY 82071)

The psychometric function in a 2IFC task, which describes the relation between the percentage of correct responses (pc) and the signal strength (x), can be expressed using the Gaussian probability function (ϕ) with a threshold (α) and a slope (β) parameter, $pc = \phi[(x/\alpha)^{\beta}/2^{1/2}]$. For auditory detection, the psychometric functions exhibited a wide range of slopes in different conditions, such as tone in noise [$\beta \approx 1$, D. M. Green, M. J. McKey, and J. C. R. Licklider, J. Acoust. Soc. Am. **31**, 1446–1452 (1959)] and tone in quiet [$\beta \approx 2$, C. S. Watson, J. R. Franks, and D. C. Hood, J. Acoust. Soc. Am. **52**, 633–643 (1972)] by listeners with acoustic hearing, and detection by listeners with electrical hearing [$\beta_{max} \approx 10$, G. S. Donaldson, N. F. Viemeister, and D. A. Nelson, J. Acoust. Soc. Am. **101**, 3706–3721 (1997)]. To understand the mechanisms underlying the change in the slope of the psychometric function, the performance of a neural detector engaged in the same tasks as the listeners was analyzed, and the neural correlates of the psychometric function, the neurometric function, was derived. Under the assumption that the detection was based on spike counts of the neuron(s) tuned to the signal frequency, the derived neurometric function showed slope values in good agreement with those obtained from the listeners.

2pPP14. Perceptual organization in loudness enhancement and decrement. Deborah A. Fantini (Dept. of Psych., Univ. of Essex, Wivenhoe Park, Colchester, CO4 3SQ, England) and Christopher J. Plack (Lancaster Univ., Lancaster, LA1 4YF, England)

Loudness enhancement is an increase in the loudness of a target tone that occurs when a more intense inducer tone either precedes or follows the target. Loudness decrement is a similar effect in the opposite direction. The present work investigates the role of perceptual organization in this paradigm by employing a capture tone presented simultaneously with the inducer. In a pair of experiments, the inducer, target, and comparison tone were always 1 kHz. The capture tone, when present, was either 2.0 or 0.5 kHz (an octave above or below the target). A 10-ms inducer tone immediately preceded a 10-ms target, with a 10-ms comparison tone presented 850 ms later. Listeners were instructed to ignore the inducer and to judge whether the target or comparison tone was louder. Equal-loudness matches were estimated using two randomly interleaved adaptive tracks, one converging on the just-louder threshold and the other on the just-quieter threshold. Results from both enhancement and decrement experiments indicate that the addition of a capture tone significantly reduces the magnitude of the effect, however it does not eliminate the effect completely. These findings indicate that perceptual organization does play a role in the effects of an inducer tone on loudness judgments.

2pPP15. Determining the area of the sweet spot in a surround loudspeaker setup for various microphone techniques. Nils Peters (Music Technol. Area, Schulich School of Music, McGill Univ., 555 Sherbrooke St. West, Montreal, PQ H3A 1E3, Canada, nils.peters@mcgill.ca) and Jonas Braasch (Rensselaer Polytechnic Inst., Troy, NY 12180-3590)

Several types of microphone techniques exist to record music performances for surround-sound reproduction. Variations between different techniques are found in the distance and angle between the microphones, and the choice of directivity patterns. All the arrays are targeted to produce an accurate spatial impression at the sweet spot. The aim of this investigation is to determine how different microphone techniques affect the size of the sweet spot, the area in which the spatial cues are reproduced with sufficient accuracy. In particular, the common belief that spaced techniques lead to larger sweet-spot areas than coincidence and near-coincident techniques is investigated. For this purpose, impulse responses (IR) of different microphone arrays are measured in a concert hall. Afterwards, test sounds are convolved with the measured IRs and presented through a surround loudspeaker setup in a control room. A dummy head is used to record the reproduced sound fields at different positions inside the listening area. In a psychoacoustic experiment, listeners are asked to rate the different recordings according to the spatial impression provided by different recording techniques. The results of the listening test will be presented and compared to signal analyzes of a binaural model.

2pPP16. Perceptual dimensions of impulsive-source active sonar echoes. Jason E. Summers, Derek Brock, Brian McClimens, Charles F. Gaumond, and Ralph N. Baer (Naval Res. Lab., Washington, DC 20375-5350)

Recent findings [J. Pitton *et al.*, J. Acoust. Soc. Am. **119**, 3395(A) (2006)] have joined anecdotal evidence to suggest that human listeners are able to discriminate target from clutter in cases for which automatic classifiers fail. To uncover the dimensions of the perceptual space in which listeners perform classification, a multidimensional scaling (MDS) experiment was performed. Subjects rated the aural similarity between ordered pairs of stimuli drawn from a set of 100 operationally measured sonar signals, comprising 50 target echoes and 50 false-alarm clutter echoes. Experimental controls were employed to evaluate consistency in judgments within and between subjects. To ensure that dimensions were discovered rather than imposed [Allen and Scollie, J. Acoust. Soc. Am. **112**, 211–218 (2002)], subjects were neither trained in classification nor made aware of the underlying two-class structure of the signal set. While training improves classification performance, prior work suggests that both expert and nave listeners use the same perceptual space [R. Mackie *et al.*, ONR Tech. Rep. 2723-1, NR 197-060/4-18-79 (1981)]. Results of the MDS study are presented, including an evaluation of the number of perceptual dimensions and their interpretation in terms of perceptual and physical attributes of the signals. [Work supported by ONR.]

2pPP17. Remote psychoacoustic testing using the Internet. Alastair C. Disley, David M. Howard, and Andrew D. Hunt (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK)

A number of recent listening tests by the authors and others have begun to use the Internet as a medium for remotely conducting psychoacoustic listening tests. While this offers a solution to the problem of gathering enough highly trained listeners in one geographic location, it results in a reduction in control of factors such as the listener's immediate environment. A recent experiment by the authors [A. C. Disley, *et al.*, "Spectral correlations of timbral adjectives used by musicians," J. Acoust. Soc. Am. **119**, 3333 abstract only (2006)] offered the opportunity to examine and quantify the effect of remote testing via the Internet. Twenty four listeners took an Internet-based listening test on ten separate occasions, alternating between a controlled environment and an uncontrolled environment. Listeners were given instructions about their uncontrolled listening environments, and data subsequently collected demonstrate good but

not universal compliance with those instructions. Comparison of experimental results between the different test environments shows very little difference, with a slight reduction in the controlled environment to the listeners' self-assessed confidence in their judgments. The techniques, advantages, and disadvantages of the Internet as a medium for remote psychoacoustic testing are discussed.

2pPP18. +Gz acceleration modifies low-frequency perception. Patrick M. B. Sandor, Lionel Pellieux (IMASSA-SC, F91223, Bretigny-sur-Orge, France), and Christophe Dehlinger (LEGTP Stanislas, F67160 Wissembourg, France)

Hypergravity is associated with maneuvering acceleration (+Gz) in fighter aircraft and can be simulated in a human centrifuge, two noisy environments where classical audiometric data in silence cannot be obtained. Yet under +4-Gz acceleration, with masking noise, the hearing threshold just superior to the masking noise upper limit frequency was reported to be slightly reduced and therefore hearing paradoxically enhanced [Sandor *et al.*, Aviat. Space Environ. Med. **75**, 9525 (2004)]. It was analyzed as an indirect evidence of attenuation of low-frequency perception in relation with the shape of the masking noise and the alteration of associated upwards spread of masking effect. A pressure modification in cochlea was supposed to occur under +Gz, due to transmission of intracranial pressure through the cochlear duct, as observed in a clinical case [Portier *et al.*, Ann. Otol. Rhinol. Laryngol. **111**, 817–820 (2002)]. We developed a finite element model of cochlear mechanics revealing a shape modification of the apical side of cochlea under the pressure modification compatible with +4-Gz acceleration. Manoussaki and colleagues [Phys. Rev. Lett. **96**(8), 088701 (2006)] emphasize the role of cochlea curvature in low frequency perception. Our experimental and computational results reinforce this role of cochlea curvature in low frequency perception.

2pPP19. Predictors of auditory performance in hearing-aid users: The role of cognitive function and auditory lifestyle. Martin D. Vestergaard[a] (Ørsted-DTU Acoust. Technol., Tech. Univ. of Denmark, DK-2800 Lyngby, and Res. Ctr. Eriksholm, Kongevejen 243, DK-3070 Snekkersten)

Within clinical audiology, it is often observed that patients who are expected to perform the same differ in auditory performance. Hearing-aid users may be dissatisfied with their instruments while they score satisfactorily in objective tests, or they may be satisfied with their instruments while no objective benefit can be measured. It has been suggested that lack of agreement between various hearing-aid outcome components can be explained by individual differences in cognitive function and auditory lifestyle. We measured speech identification, self-report outcome, spectral and temporal resolution of hearing, cognitive skills, and auditory lifestyle in 25 new hearing-aid users. The purpose was to assess the predictive power of the nonauditory measures while looking at the relationships between measures from various auditory-performance domains. The results showed that only moderate correlation exists between objective and subjective hearing-aid outcome. Different self-report outcome measures showed a different amount of correlation with objective auditory performance. Cognitive skills were found to play a role in explaining speech performance and spectral and temporal abilities, and auditory lifestyle was correlated with self-report outcome. However, overall the predictive leverage of the various measures was moderate, with single predictors explaining only up to 19 percent of the variance in the auditory-performance measures. [a] Now at CNBH, Department of Physiology, Development and Neuroscience, University of Cambridge, UK.

2pPP20. Perception and production of virtual pitch by absolute-pitch musicians. I-Hui Hsieh and Kourosh Saberi (Dept. of Cognit. Sci., Univ. of California at Irvine., Irvine, CA 92697, ihsieh@uci.edu)

The ability of absolute pitch (AP) and control (non-AP) musicians to identify and produce musical pitch from long-term memory was investigated in three experiments. In experiment 1, subjects identified the pitch of either a missing-fundamental complex or a highpass-filtered piano note. Unlike control subjects, AP subjects were highly accurate (¿85%) in identifying virtual pitch even when cues were restricted to the 4th and 5th harmonics. In a second experiment, subjects produced, within 3 s, the pitch of a randomly selected musical note without an external acoustic reference through either (1) adjusting a octave-range slider or (2) voicing/singing the target note. AP subjects, but not controls, were highly accurate (0.3 semitone) in the adjustment task. However, both AP and non-AP subjects were extremely accurate (i.e., SD=0.97, 0.90 semitone error, respectively) in voicing the required pitches as determined from FFTs of the recorded voiced notes. In experiment 3, subjects identified pitch of solfeggio syllables at mismatched note frequencies, created by altering the sampling rate of digitally recorded voiced labels. Identification accuracy by AP subjects declined to 65%, supporting a cognitive/linguistic interference model. Implications of these findings for models of absolute pitch encoding will be discussed. [Work supported by NSF.]

2pPP21. Pitch estimation of bone-conducted ultrasound. Kazuhito Ito and Seiji Nakagawa (Natl. Inst. of Adv. Industrial Sci. and Technol., 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan, keith.ito@aist.go.jp)

Although ultrasound is perceptible through bone conduction, its perception mechanism has been clarified only slightly. The subjective pitch of bone-conducted ultrasound (BCU) sounds like a tone or harmonic sound of ten-odd kilohertz, independent of its own frequency. It has been difficult to determine the frequency of the pitch precisely because the pitch varies among listeners and with the position and strength of BCU stimulation, moreover; BCU hearing is not masked well by air-conducted (AC) band-pass noise. In this study, the BCU pitch-perception mechanism was investigated using beats with AC tones. The AC tones of 10–17 kHz, for which the presence of the pitch of BCU had been expected, were presented sequentially together with BCU (30 kHz) stimulation. Listeners were asked whether or not they were able to hear beats. For comparison, bone-conducted (BC) harmonic sounds of audible frequency components (e.g., 11–16 kHz, 1-kHz step) were also investigated. Results showed that no beat was heard between the AC tones and the BCU, whereas beats between the AC tones and the BC harmonic sounds were heard, suggesting the possibility that the pitch-perception mechanism of BCU is different from those of AC and BC sounds in this audible frequency area.

2pPP22. Off-frequency listening in a pitch matching experiment. Veronica Eckstein and Bruce G. Berg (Dept. of Cognit. Sci., Univ. of California Irvine, 3151 Social Sci. Plaza, Irvine, CA 92697)

A pitch-matching paradigm was used to investigate the pitch of three-tone complexes. For different conditions, the central tone was 200, 1000, or 2000 Hz with the side tones spaced by either 30 or 45 Hz. The tones were of equal intensity or a signal was added in phase to the central tone with a level of 5, 0, or 5 dB relative to the side tones. Listeners adjusted the frequency of a pure tone to match the pitch of the complex. According to a model known as the envelope weighted average of the instantaneous frequency (EWAIF) [L. Feth, Percept. Psychophys. **15**, 375–378 (1974)], the pitch should be approximately equal to the center frequency for all conditions. Contrary to this prediction, increasing the bandwidth or decreasing the signal level systematically lowered the pitch. However, the results are consistent with EWAIF calculations if the stimuli are first passed through the skirt of an auditory filter. This method may provide an alternative technique for estimating the filtering properties of the auditory system. [Work supported by Center for Hearing Research, UCI.]

2pPP23. Sample discrimination of frequency differences with ipsilateral and contralateral distracters. Donna L. Neff, Walt Jesteadt, and Lori J. Leibold (Boys Town Natl. Res. Hospital, 555 N 30th St., Omaha, NE 68131)

This experiment examined the effect of ipsilateral and contralateral tonal distracters on sample discrimination of frequency differences. In the 2AFC baseline task, four listeners indicated which of two intervals, each containing a two-tone target complex, was higher in pitch. The pairs of target tones were drawn from Gaussian frequency distributions centered at low-, middle-, and high-frequency regions. Two distracter tones were then added, one from each of the two nontarget frequency regions (e.g., low targets, middle and high distracters). Distracter tones were either randomly drawn from a Gaussian frequency distribution or fixed in frequency at the mean of the distribution. Across conditions, targets were in the left or right ears, with distracters in the ipsilateral or contralateral ear. There was little effect of target ear or effect of adding fixed-frequency distracters in either ear. Effects of random-frequency distracters varied across subjects and frequency regions, with both ipsilateral and contralateral distracters showing large effects for some subjects and conditions. Distracters were least effective with middle-frequency targets. The ability to use the ear as a channel appears to depend on both the relative variability and relative frequency regions of the targets compared to the distracters. [Work supported by NIDCD.]

2pPP24. Dynamic center of gravity effects: Matching virtual frequency glides. Evelyn M. Hoglund, Lawrence L. Feth (Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210), and Ashok K. Krishnamurthy (The Ohio State Univ., Columbus, OH 43210)

Lublinskaja (1996) demonstrated that changing the amplitudes of two-resonance Russian vowels over time leads listeners to hear a glidelike transition that follows the dynamic center of gravity (COG). Ananthara-man (1998) investigated COG transitions for sinusoidal signals by asking subjects to match frequency-modulated (FM) to virtual frequency (VF) glides. The amplitudes of two tones were simultaneously modulated to produce a percept of a glide with rising pitch, similar to that of an FM sweep, although the frequencies remained fixed. However, listeners might perform Anantharaman's task by simply attending to the location of one of the anchor tones without perceiving the virtual glide, as the frequencies of the two component tones were changed to alter the frequency span of the VF glide. This study modified that work by fixing the anchor frequencies, but confirmed the potential for procedural artifacts. Two additional experiments eliminated FM signals altogether and asked listeners to match VF glides to VF glides. Even with anchor frequency roving, subjects were able to match the VF glide trajectories for frequency separations up to 6 to 8 ERB. Results were compared with predictions from the perceptual spectral centroid model. [Research supported by a grant from NIH/NIDCD R01-DC006879.]

2pPP25. Spectrotemporal structure influences spatial release from informational masking in reverberant settings. Erol Ozmeral, Priyanka Zutshi, Virginia Best, and Barbara Shinn-Cunningham (Dept. of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215, ozmeral@cns.bu.edu)

Spatial separation between target and masking sources improves target identification due to both changes in energy (acoustic benefit) and the ability to focus attention on the target (spatial benefit). Studies of speech intelligibility show that while acoustic benefits are reduced by reverberation, spatial benefits remain robust [Kidd *et al.*, Acustica **114**, 526–536 (2005)]. The current study extends a previous study of nonspeech communication calls [birdsong; Best *et al.*, J. Acoust. Soc. Am **118**, 3766–3773 (2005)] to explore whether, like speech, the spatial benefit for these signals is robust in reverberation. It was found, however, that reverberation decreased both acoustic and spatial benefits for birdsong identification. To examine whether the decrease in the spatial benefit was due to spectrotem-

poral differences between speech and birdsong, songs were slowed down to half the original rate (making them more similar to speech). For these songs, the spatial benefit was preserved in reverberation. These results suggest that spatial benefits in reverberation are less effective for rapidly fluctuating signals, presumably because reverberation smears their temporal structure, impairing segregation of competing syllables. Thus, the relatively slow fluctuations in speech may be important for its resilience to interference from other talkers in reverberant environments. [Work support by Grants AFOSR-FA9550-04-1-0260 and NIDCD (P30) DC04663.]

2pPP26. Effect of number of masking talkers on masking of Chinese speech.

Xihong Wu, Jing Chen, Zhigang Yang, and Liang Li (Speech and Hearing Res. Ctr., Peking Univ., Beijing, China, 100871, wxh@cis.pku.edu.cn)

In this study, speech targets were nonsense sentences spoken by a Chinese female, and speech maskers were nonsense sentences spoken by other one, two, three, or four Chinese females. All stimuli were presented by two spatially separated loudspeakers. Using the precedence effect, manipulation of the delay between the two loudspeakers for the masker determined whether the target and masker were perceived as coming from the same or different locations. The results show that the one-talker masker produced the lowest masking effect on Chinese speech. When the number of masking talkers increased progressively to 2, 3, and 4, even though informational masking progressively decreased, energetic masking progressively increased, leading to an increased total masking effect on targets. A new form of calculation of the speech intelligibility index confirmed an increase of energetic masking as the masking-talker number increased, even when the long-term average signal-to-noise ratio was unchanged. Some differences between Chinese speech masking and English speech masking were revealed by this study. [Work supported by China NSF and Canadian IHR.]

2pPP27. The role of fundamental frequency and target segregation in the presence of same-band and different-band maskers.

Nandini Iyer, Douglas Brungart, and Brian Simpson (Air Force Res. Lab., 2610 Seventh St., Bldg 441, Area B, WPAFB, OH 45431, nandini.iyer@wpafb.af.mil)

A number of recent studies on multitalker speech perception have been based on envelope-modulated, sine-wave-excited speech similar to the kind used to simulate the processing used in cochlear implants. Although these modulated sine-band signals are highly intelligible in isolation, they are typically much harder to segregate in multitalker listening tasks than normal speech. In this experiment, we examined the extent to which the difficulty in segregating these signals could be attributed to the absence of the fundamental frequency $F0$ that is present in natural speech. Psychometric functions were measured with speech and noise maskers for three types of signals: (1) natural speech; (2) sine-excited cochlear implant speech (CIS); and (3) $F0$-excited CIS, which was generated in the same way as the CIS but with a simulated glottal pulse train as an excitation source. The results show that performance was better with the $F0$-excited CIS than with the sine-excited CIS in all conditions, including those where the target and masking talkers overlapped in frequency. Performance was also generally better when the $F0$-excited CIS was generated with $F0$ values from different-sex rather than same-sex talkers. The results confirm the importance of $F0$ cues in the segregation of competing speech signals.

2pPP28. Effect of peripheral interference and attentional loss on divided listening to multiple simultaneous talkers.

Takayuki Kawashima (Dept. of Cognit. and Behavioral Sci., Univ. of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, 153-8902 Japan, ckawa@mail.ecc.u-tokyo.ac.jp) and Takao Sato (Univ. of Tokyo, Bunkyo-ku, Tokyo, 113-0033 Japan)

Effects of two factors, peripheral interference (masking) and attentional loss, that may impose restriction on divided listening were investigated using Japanese words. In each trial in the divided listening task, simultaneous multiple voices were presented first, then followed by a target voice produced by one speaker. Participants reported whether the target was included in the multiple voices or not. When the number of talkers (manipulated from one to six) was more than three, the estimated hit ratio was lower than unity, and decreased monotonically as the number of talkers increased. To investigate factors inducing the decrement, participants were also tested when the target was presented before the multiple voices. In this condition, decrement of the performance as a function of the number of talkers was gradual, compared to that in the divided listening condition. Decrement of the performance in the pretarget condition was thought to come mainly from peripheral interference, and the difference between the two tested conditions was interpreted as representing the cost in dividing attention. The result suggests that failure in divided listening to three or four talkers largely comes from the attentional factor, although the role of peripheral interference becomes larger when there are more talkers.

2pPP29. Change deafness and how the auditory system organizes sounds.

Melissa K. Gregg and Arthur G. Samuel (Dept. of Psych., State Univ. of New York, Stony Brook, Stony Brook, NY 11794-2500, mgregg@ic.sunysb.edu)

Change blindness, or the failure to detect changes to a visual scene that seem blatantly obvious after the fact, has been demonstrated in a variety of different situations. Exploration of auditory failures to detect change has been far less studied, even though demonstrations of change deafness could have important theoretical implications. Four experiments were conducted to explore the processes involved in change deafness by measuring explicit change detection as well as auditory object encoding of natural auditory scenes. Considerable change deafness occurred, even though auditory objects were encoded quite well. Change detection performance was better when stimuli were not presented at distinct spatial locations (experiment 1) than when they were presented with spatial cues (experiment 2). Familiarity with the objects did not affect detection or recognition performance. In order to establish an effective cue for auditory detection and recognition, performance was assessed when auditory stimuli were presented in distinct regions of perceptual space based on fundamental frequency and harmonicity (experiments 3 and 4). Implications for the mechanisms responsible for change deafness and auditory sound organization are discussed.

2pPP30. Sensitivity to increments analyzed separately for each level in a roving paradigm.

Anne-Marie Bonnel and Ervin Hafter (Dept of Psych., Univ. of California at Berkeley, 3210 Tolman Hall, Berkeley, CA 94720-1650)

Studies of level discrimination using a dual-task paradigm have shown no cost of shared attention when levels were roved, so that comparisons between targets and standards were in sensory-trace memory, but clear evidence of a cost when levels were fixed, and comparisons seemed based on context-coded memory. Interest in the effects of these two types of memory led here to an auditory single-task in which a trial consisted of a standard tone set to level (I), followed by a target that was (I) alone or (I+ΔI). Durations were 50 or 200 ms; standards ranged from 50 to 70 dB SPL. In line with the view that the near miss to Weber's Law is negligible over this range, fixed standards provided constant thresholds (log ΔI/I). As typical, overall performance was reduced by roving, but a surprising result was that the magnitude of this loss depended on (I). Thus, roving had the least effect with the highest value of (I); then thresholds rose steeply as (I) was reduced. Independent of this slope, a wider range of roving produced a greater overall loss. This is discussed in terms of an additive noise, possibly connected to the use of trace memory.

2pPP31. Factors influencing the auditory and cross-modal attentional blink. Adelbert W. Bronkhorst, Marieke van der Hoeven (TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands, adelbert.bronkhorst@tno.nl), Jan Theeuwes, Erik van der Burg, and Thomas Koelewijn (Vrije Univ., 1081 BT Amsterdam, The Netherlands)

When an observer has to report two targets (T1 and T2) embedded in a stream of distracters, an attentional blink can occur. This is the degradation in performance for T2 when T1 and T2 are presented sufficiently close in time (within about 500 ms). The attentional blink paradigm has been extensively used within the visual domain as a tool to study bottlenecks in the processing of sensory information. Several studies have shown that an attentional blink can also occur during auditory or cross-modal presentation. However, these blinks are smaller and more stimulus dependent than visual blinks. In our research we have found a robust blink when T1 and T2 are tones embedded in distracters tones, but no blink when T2 is replaced by a visually presented letter among different distractor letters. The cross-modal blink is also absent when both T1 and T2 are digits. We do find a clear cross-modal blink when a visual T1 with increased difficulty (three letters or a word) is combined with a T2 consisting of a spoken letter. Our results indicate that there are modality-specific as well as amodal bottlenecks and that the latter occur when stimuli are processed on a verbal or semantic level.

2pPP32. Reproducible maskers reveal similar mechanisms predicting multiple-burst-same and multiple-burst-different informational masking conditions. Robert H. Gilkey (Wright State Univ., Dayton, OH 45435), Christine R. Mason, and Gerald Kidd, Jr. (Boston Univ., Boston, MA 02215)

Gilkey, Mason, and Kidd [J. Acoust. Soc. Am. **109**, 2468 (2001)] found that although informational masking has typically been thought to depend more strongly on the variability in the ensemble of masking stimuli presented across trials, trial-by-trial performance was strongly related to the specific masker waveform presented on each trial. Moreover, hit and false-alarm rates to individual masker stimuli under multiple-burst-same (MBS) conditions were well predicted by the patterns of energy in those stimuli. Hit and false-alarm rates under the multiple-burst-different (MBD) condition were more difficult to predict. This paper reports on further efforts to model those data. The results indicate that performance under both MBS and MBD conditions can be predicted using a linear combination of envelope statistics computed from seven narrow frequency bands surrounding the 1000-Hz signal frequency. MBS responses are related to the average height of the envelope in these bands and MBD responses are related to the standard deviations of the envelopes. [Work supported by NIH/NIDCD.]

2pPP33. The application of the equal energy hypothesis (EEH) to interrupted, intermittent, and time-varying non-Gaussian noise exposures. Roger P. Hamernik, Wei Qiu, and Robert I. Davis (Auditory Res. Lab., State Univ. of New York at Plattsburgh, 101 Broad St., Plattsburgh, NY 12901)

Industrial data and animal research show that non-Gaussian (nonG) noise exposures are more hazardous to hearing than energy equivalent Gaussian (G) exposures. A statistical metric, kurtosis $[b(t)]$, was shown to order the severity of noise-induced trauma following a nonG exposure relative to an energy and spectrally equivalent G exposure. Four groups of chinchillas were exposed to one of four different nonG interrupted, intermittent, and time-varying (IITV) noise paradigms over 19 days at an Leq= 103 dB(A) SPL, with $b(t)$= 25 or 50. Each daily exposure consisted of two 4.25-h periods with an hour break. Each 4.25-h exposure was interrupted for 15 min and each 5-day sequence was separated by a 2-day break. Each daily IITV exposure followed one of two different SPL temporal patterns that varied between 90 and 108 dB(A). All IITV exposures produced a toughening effect that did not alter the degree of noise-induced trauma. NonG noise produced as much trauma as a G exposure at 110 dB.

Despite very different temporal paterrns for the $b(t)$= 50 exposures, trauma was the same. Thus within a common class [i.e., the same Leq and $b(t)$] of nonG, IITV exposure, the EEH may apply. [Work supported by NIOSH.]

2pPP34. The application of statistical learning models to the prediction of noise-induced hearing loss. Wei Qiu, Jun Ye, and Roger P. Hamernik (Auditory Res. Lab., State Univ. of New York at Plattsburgh, 101 Broad St., Plattsburgh, NY 12901)

Three powerful nonlinear statistical algorithms [a support vector machine (SVM), radial basis function network (RBFN), and regression tree] were used to build prediction models for noise-induced hearing loss (NIHL). The models were developed from an animal (chinchilla) database consisting of 322 animals exposed to 30 Gaussian and non-Gaussian noise conditions. The inputs for the models were either energy or energy plus kurtosis. The models predict inner hair cell (IHC) loss, outer hair cell (OHC) loss, and postexposure threshold shift (PTS) at 0.5, 1, 2, 4, and 8 kHz. The models incorporating both energy and kurtosis improved the prediction performance significantly. The average performance improvement for the prediction of IHC loss was as much as 55%, for OHC loss it was 66% and for PTS, 61%. The prediction accuracy of SVM and RBFN with energy plus kurtosis for all three outputs (predictions) was more than 90% while for the regression tree model it was more than 85%. Energy is not a sufficient metric to predict hearing trauma from complex (non-Gaussian) noise exposure. A kurtosis metric may be necessary for the prediction of NIHL. [Research supported by NIOSH.]

2pPP35. The development of models for the prediction of noise-induced hearing loss. Wei Qiu, Jun Ye, and Roger P. Hamernik (Auditory Res. Lab., State Univ. of New York at Plattsburgh, 101 Broad St., Plattsburgh, NY 12901)

Three statistical learning models were developed to predict noise-induced hearing loss (NIHL) from an archive of animal noise exposure data, which contains 936 chinchillas exposed to various noise environments. The following models were constructed: (*i*) A support vector machine model with a nonlinear radial basis function kernel. (*ii*) A multilayer perceptron network model and (*iii*) a radial basis function network model. In addition to frequency-specific energy metrics, noise exposure parameters and biological metrics such as kurtosis, noise type, and pre/postexposure hearing thresholds were used as inputs to the model. There were several indices of auditory trauma at specific audiometric test frequencies that were to be predicted by the models: e.g., noise-induced permanent threshold shift, percent outer hair cell loss, and percent inner hair cell loss. The average prediction accuracy for the three models was better than 80%. These results demonstrate the feasibility of developing such models for the prediction of NIHL in humans. [Research supported by NIOSH.]

2pPP36. Noise-induced hearing loss from non-Gaussian equal energy exposures. Robert I. Davis, Wei Qiu, and Roger P. Hamernik (Auditory Res. Lab., State Univ. of New York at Plattsburgh, 101 Broad St., Plattsburgh, NY 12901)

Data from several different exposures [Hamernik *et al.*, J. Acoust. Soc. Am. **114**, 386–395 (2003)] showed that, for equivalent energy [Leq = 100 dB(A)] and spectra, exposure to a continuous, non-Gaussian (nonG) noise produces greater hearing and sensory cell loss in the chinchilla than a Gaussian (G) noise. The statistical metric, kurtosis, could order the extent of the trauma. We extend these results to Leq= 90 and 110 dB(A), non-Gaussian noises generated using broadband noise bursts, and band-limited impacts within a continuous G background noise. Data from nine new experimental groups with 11 or 12 chinchillas/group will be presented. Evoked response audiometry established hearing thresholds and surface preparation histology quantified sensory cell loss. There were clear

intensity-related effects. At the lowest levels there were no differences in the trauma produced by G and nonG exposures. At Leq=90 dB(A), nonG exposures produced increased trauma relative to equivalent G exposures. By removing energy from the impulsive transients by limiting their bandwidth, trauma could be reduced. The use of noise bursts to produce the nonG noise also reduced the amount of trauma. A metric based on kurtosis and energy may be useful in modifying existing exposure criteria. [Work supported by NIOSH.]

2pPP37. The application of the equal energy hypothesis (EEH) to interrupted, intermittent and time-varying Gaussian noise exposures. Wei Qiu, Roger P. Hamernik, and Robert I. Davis (Auditory Res. Lab., State Univ. of New York at Plattsburgh, 101 Broad St., Plattsburgh, NY 12901)

Interrupted, intermittent, time-varying (IITV) noise exposures may reduce the susceptibility of the cochlea to hearing loss through toughening or conditioning effects. Six groups of chinchillas were exposed to a broadband Gaussian IITV noise over the course of 19 days at an Leq=101 or 106 dB(A) SPL. All exposures at a given Leq had the same total energy. Each daily exposure consisted of two 4.25-h periods with an hour break in between. Each 4.25-h exposure was interrupted for 15 min and each 5-day sequence was separated by a 2-day break. Daily IITV exposures at 101 dB varied between 90 and 108 dB(A), while those at 106 dB(A) varied between 80 and 115 dB(A). For the IITV exposures three different variations in the SPL temporal profile were used. Five-day continuous, equal energy exposures at 100 or 106 dB(A) SPL served as control conditions. The IITV exposures produced up to 40 dB toughening that did not have any effect on thresholds or sensory cell losses. There were some differences in the permanent threshold shift and cell loss across equal energy exposures but the differences were small. The EEH may be valid for Gaussian IITV exposures. [Work supported by NIOSH.]

2pPP38. Temporal characteristics of extraction of size information in speech sounds. Chihiro Takeshima (Grad. School of Music, Kyoto City Univ. of Arts, Kyoto 610-1197, Japan, ctakeshima@yahoo.co.jp), Minoru Tsuzaki (Kyoto City Univ. of Arts, Kyoto 610-1197, Japan), and Toshio Irino (Wakayama Univ., Wakayama 640-8510 Japan)

We can identify vowels pronounced by speakers with any size vocal tract. Together, we can discriminate the different sizes of vocal tracts. To simulate these abilities, a computational model has been proposed in which size information is extracted and separated from the shape information. It is important to investigate temporal characteristics of the size extraction process. Experiments were performed for listeners to detect the size modulation in vowel sequences. All the sequences had six segments. Each segment contained one of three Japanese vowels: "a," "i," and "u." Size modulation was applied by dilating or compressing the frequency axis of continuous, STRAIGHT spectra. Modulation was achieved by changing the dilation/compression factor in sinusoidal functions. The original $F0$ pattern of the base sequence, except for warping of the time axis, was used for all stimuli. The minimum modulation depth at which listeners were able to detect the existence of modulation was measured as a function of the modulation frequency. The results will be compared with low-pass characteristics in a temporal modulation transfer function obtained with the amplitude-modulated noise. They will be discussed in relation to a computational model based on the Mellin transformation.

2pPP39. She hears seashells: Detection of small resonant cavities via ambient sound. Ethan J. Chamberlain, Lawrence D. Rosenblum, and Ryan L. Robart (Dept. of Psych., Univ. of California, Riverside, Riverside, CA 92521)

There is evidence that blind listeners can detect the presence of obstacles based on how the obstacles structure the ambient sound in a quiet room [Ashmead *et al.*, J. Vis. Impair. Blindn. **9**, 615 (1998)]. Potentially,

this ability is based on auditory sensitivity to the buildup of ambient wave interference patterns in front of the obstacle. This effect is not unlike how the internal structure of a seashell amplifies a room's ambient sound to emulate the sound of the ocean. Experiments were conducted to determine whether unpracticed, sighted listeners were sensitive to this information. Blindfolded listeners were placed in a quiet room and were asked to determine the location of a small resonant cavity placed next to either their left or right ear. Results revealed that listeners were very accurate at this task. Follow-up experiments examined the limits of this sensitivity as well as its acoustical support. These findings suggest that unpracticed listeners might be sensitive to subtle changes in the ambient acoustic structure of a quiet environment.

2pPP40. Effects of listening environment on speech recognition in noise. Linda Thibodeau and Tina Keene (Univ. of Texas at Dallas, 1966 Inwood Rd., Dallas, TX 75235)

The ability to recognize speech in noisy situations varies among individuals despite normal hearing. Although some of the variation may be related to physiological differences, some may be attributed to environmental experiences. The purpose of this investigation was to examine how everyday listening experiences were related to speech recognition in noise abilities. Using the Hearing in Noise Test, reception thresholds for speech were compared for listener groups who differed in their preferred listening levels for music and signal-to-noise ratios for daily communication. Results suggested that speech recognition in noise was related to listening levels. These results have implications for the use of assistive devices to improve the signal-to-noise ratio in adverse communication situations.

2pPP41. Comparing vowel formant thresholds from two tasks: Classification versus 2-alternative forced choice (2AFC) adaptive tracking. Eric Oglesbee (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405, eoglesbe@indiana.edu) and Diane Kewley-Port (Indiana Univ., Bloomington, IN 47405)

Accurate classification of vowels in American English is difficult because of the number of acoustically similar vowels. Previous experiments in our lab [Kewley-Port *et al.*, J. Acoust. Soc. Am. **118**, 1929–1930 (2005)] described a new method to obtain formant thresholds using a classification task rather than the more typical 2AFC task. The goal of the present experiment was to directly compare formant thresholds estimated from a classification task with a modified 2AFC paradigm using adaptive tracking. Stimuli, generated by STRAIGHT, consisted of separate continua for formants F1 and F2 based on natural productions of "bid" shifted to "bed," and "cut" shifted to "cot." Separate 7- and 14-step continua were created for the classification and 2AFC tasks, respectively. Eight subjects participated first in the classification task, followed by the 2AFC task. Classification threshold results here replicated those observed in our previous experiments. Within-subject comparisons of the classification and 2AFC data refined our method for calculating a classification threshold from logistic psychometric functions. Individual thresholds calculated using this refined method were very similar to those from the 2AFC task. Results demonstrate the viability of using a more natural, single-interval classification task to estimate formant discrimination thresholds. [Work supported by Grant NIHDCD-02229.]

2pPP42. Cochlear and cognitive models for speech enhancement. Melissa Dominguez (DBK Acoust., 110 Village Station Ln., Grayslake, IL, 60030), Jeff Bondy, and Andrew Dittberner (GN Resound, Glenview, IL, 60026)

There has been a long running interest in using cochlear models for audio processing. Psychophysical masking stemming from normal cochlear operation is used as the basis for most audio compression circuits, while the cochleas mel-frequency response is mimicked by most automatic speech recognition circuits. An area that has been less well explored

in audio processing is the auditory brain. Recent research shows that cognitive abilities are a good predictor of peformance on many hearing tasks, especially those which are more complex and realistic. By extending a nonlinear, adaptive cochlear model with cognitive modules a novel tool for sound assessment is made that allows an estimate of cognitive load. Quantitative offline benefit has never been modeled before. We test the new predictor on noise reduction circuits in hearing aids. This family of algorithms is well known to not increase intelligibility but has been shown to reduce perceived noisiness. This reduction is proposed as equivalent to reducing cognitive load. The elderly, with their heavier reliance on high level processing, could greatly benefit from the reduction in cognitive load. A new algorithm for reducing cognitive load is compared to spectral subtraction algorithms.

2pPP43. Cross-modal plasticity and the sensitive period for cochlear implantation. Kristi A. Buckley, Emily A. Tobey (The Univ. of Texas at Dallas, 1966 Inwood Rd. Dallas, TX 75235), and Anu Sharma (The Univ. of Colorado at Boulder)

Previous research suggests prelingually deaf individuals who receive a cochlear implant early in childhood develop better oral communication skills than those who receive their implant later. However, the mechanisms influencing this sensitive period are not well understood. This investigation explores one possible mechanism limiting the sensitive period for cochlear implantation, cross-modal plasticity. Cross-modal plasticity occurs when one sensory system recruits cortical processing resources of another absent or damaged sensory system. This reallocation of cortical processing resources may influence the ability of the auditory system to respond to restored auditory sensation through the electrical stimulation of a cochlear implant. In order to investigate the possible relationship between cross-modal plasticity and the sensitive period for cochlear implantation, we examined the visual evoked potential (VEP) to peripheral motion in prelingually deaf children and adults, and normal-hearing children and adults. The VEP response source is compared across ages and groups. Preliminary data from adults indicate the source of the VEP response in deaf individuals originates from the temporal lobe, compared to the parietal lobe in normal hearing individuals, suggesting evidence of cortical reorganization in deafness. [Work funded by NIH F31DC007007 (K. Buckley, P.I.), R01DC04558 (E. Tobey, P.I.), and R01DC006257 (A. Sharma P.I.).]

2pPP44. Predictions of Mandarin Chinese tone identification in cochlear implants using particle filters. Kenneth D. Morton, Jr., and Leslie M. Collins (Dept. of Elec. and Comput. Eng., Duke Univ., 130 Hudson Hall, Box 90291, Duke Univ., NC 27708)

The spectral information provided in current cochlear implants is inadequate for reliable speech perception in tonal languages. Speech comprehension in a tonal language, such as Mandarin Chinese, requires recognition of lexical tones. New strategies for sound processing may allow the transmission of more spectral information and may therefore improve Mandarin Chinese tone recognition. Four cochlear implant speech processing strategies were examined in this study. Each provides a different amount of spectral information. Automated pattern recognition of Mandarin Chinese tones processed using cochlear implant acoustic models can be used as a means of testing the abilities of each of the algorithms to transmit the necessary spectral information for accurate tone perception and may be used to focus future Mandarin Chinese tone recognition studies in cochlear implant subjects on the most promising speech processing strategies. Classification based on spectral features derived from particle filtering is compared to classification derived from other features in an effort to determine the necessary spectral information for accurate tone classification in both quiet and noisy conditions. Automated classification results will be compared to the results of a listening experiment using normal-hearing subjects and acoustic models.

2pPP45. Simulating the effect binaural (electric and acoustic) hearing on Mandarin speech recognition under background noise for patients with cochlear implant (CI). Chao-Min Wu and Wei Jen Cheng (Dept. of Elec. Engr., Natl. Central Univ., 300, Chung-Da Rd., Chung-Li, Taiwan 320)

This study evaluated the effects of binaurally combined electric and acoustic hearing on Mandarin speech recognition under simulated background noise for normal-hearing (NH) listeners and CI recipients of the Nucleus-24 device with contralateral hearing aid (HA). Chinese word and tone recognition were measured in seven NH subjects using simulated four-channel continuous interleaved sampler (CIS) type speech processing strategy under three different conditions, namely, (1) only simulated (CI) signals with modulated Gaussian white noise for monaural stimulation, (2) simulated (CI) signals with modulated Gaussian white noise and simulated HA signals (low-pass speech signals with cutoff frequency of 500 Hz) for monaural stimulation, and (3) simulated (CI) signals with modulated Gaussian white noise to one ear and simulated hearing aid (HA) signals to the other ear for dichotic stimulation. For comparison purposes, Chinese word and tone recognition were also measured in three CI users using advanced combination encoding (ACE) strategy. Preliminary results showed that the Chinese word and tone recognition rates of the NH listeners and CI recipients are statistically significantly improved (word and tone recognition rates were improved from 3.85% to 60.86% and 40.14% to 97.43% for NH listeners and 20% to 37.5% and 50% to 90% for CI recipients).

2pPP46. Pitch discrimination in children using cochlear implants. Amy McKinnon and Michael Kiefte (School of Human Commun. Disord., Dalhousie Univ., 5599 Fenwick St., Halifax, NS, B3H 1R2 Canada)

Research in pitch perception by cochlear implant users has focused primarily on postlingually deafened adults. Little is currently known regarding the pitch perception abilities of those who are implanted at a very early age, yet many children receive cochlear implants when they are less than 2 years old. Given the effects of brain plasticity, it is conceivable that pitch perception by early cochlear implant users may be much better than that of their older counterparts. This study examines pitch discrimination thresholds in young cochlear-implant users between the ages of 4 and 16. A computer game was developed to determine difference limens for fundamental frequency in vowellike stimuli at three referent frequencies: 100, 200, and 400 Hz. Two different tasks were employed: pitchdiscrimination, in which subjects were asked whether two stimuli are same or different, and pitch ranking, in which participants determine which of two stimuli is higher in pitch. Normal-hearing matches were tested to control for maturational effects, and adult cochlear implant users were also tested to compare findings with other studies. [Work supported by the Nova Scotia Health Research Foundation.]

2pPP47. Improving the voice pitch discrimination threshold through cochlear implants. Shizuo Hiki (Waseda Univ., Tokorozawa, Japan, hiki@waseda.jp) and Masae Shiroma (Natl. Univ. of Health and Welfare, Ohtawara, Japan)

Aiming at the maximum use of linguistic information conveyed by the periodicity pitch of speech sound through cochlear implants, the possibility of improving the voice pitch discrimination threshold was examined for cochlear implantees: three acquired deaf adults and four congenitally deaf children. Female and male adult speech samples of vowel /a/ were analyzed and resynthesized using the STRAIGHT computer program to shift the original voice pitch (c1 and c) one octave upward and downward with a half-step interval on the musical scale. By presenting those synthetic speech samples in pairs, the discrimination threshold of the voice pitch interval was observed. Association of the speech samples to the picture of a piano keyboard was helpful for conceptualizing the voice pitch change, especially for children. Preliminary observation results suggested that the discrimination threshold can be improved by less than one

whole step by adults and two whole steps by children. Those figures were promising because, even for Japanese word accent and Chinese tones, which were conveyed by the precise changes in voice pitch, it was ascertained by acoustical analyses of both their production and perception that discrimination of a half-step difference was small enough to recognize their pitch patterns.

2pPP48. Implementing a psychophysical interface for the sound processor for electrical and acoustic research version 3 (SPEAR3). Joshua S. Stohl, Chandra S. Throckmorton, and Leslie Collins (Duke Univ., 130 Hudson Hall, Box 90291, Durham, NC 27708; lcollins@ee.duke.edu)

Current speech processing strategies use amplitude-modulated fixed rate pulse trains to communicate information. Using acoustic models, researchers have recently demonstrated that variable stimulation rate strategies may provide an improvement of speech perception in noise and music perception [Nie *et al.*, 2005; Stohl and Collins, 2006; Throckmorton *et al.* (2006) (in press)]. It has been hypothesized that successful implementation of these strategies will require psychophysics-based tuning. An experimental interface is needed to implement such psychophysical experiments, and most interfaces provided by implant manufacturers do not provide the stimulus design flexibility necessary. HearWorks provides the SPEAR3 developmental sound processor, which is capable of dynamically updating the stimulation rate and offers bimodal and bilateral configurations. Assembly and Visual Basic code is included with the SPEAR3, but must be modified and expanded to implement dynamic stimulus presentation and graphical user interfaces for specific psychophysical tasks. Using data collected from Nucleus CI 24 implant users, we will demonstrate that the SPEAR3 may be used to implement general psychophysical experiments. The ability of the SPEAR3 speech processor to implement variable stimulation rates is critical to future psychophysical testing of variable stimulation rate processing strategies.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006

REGENCY BALLROOM (ROYAL HAWAIIAN), 2:00 TO 4:45 P.M.

Session 2pSA

Structural Acoustics and Vibration: Scattering, Diffraction, Radiation, and Damping

Sabih I. Hayek, Cochair
Pennsylvania State Univ., Dept. of Engineering Science and Mechanics, 212 Earth Engineering Science Bldg., State College, PA 16802-6812

Katsuji Akamatsu, Cochair
Machinery Acoustics, 1-1-2-314 Obanoyama, Shinohara, Nada-ku, Kobe 657-0015, Japan

Contributed Papers

2:00

2pSA1. Phased array steering of structure-borne sound radiated from a flexible beam by active modal control. Alfred Tan (Dept. of Aerosp. Eng., Tokyo Metropolitan Univ., 6-6 Asahigaoka, Hino-shi, Tokyo, 191-0065, Japan), Dunant Halim (The Univ. of Adelaide, SA 5005, Australia), and Nobuo Tanaka (Tokyo Metropolitan Univ., Hino-shi, Tokyo, 191-0065, Japan)

This paper analytically demonstrates the implementation of a control strategy to actively steer the structure-borne sound radiated from a simply supported beam structure, which is excited by an arbitrary mono-frequency force. The surface of the beam is discretized into many finite elemental sound radiators that correspond to their respective nodal points. The dynamics of the beam structure are identified in order to modify the modal vibration characteristic for beamsteering purpose. The acoustic power is formulated in terms of its velocity according to the Rayleigh expression. The key to the control strategy is to combine the *a priori* knowledge of the structural modal characteristic with the sensors measurements. As a result, this is used to control the normal surface vibration velocity profile of the beam structure into a set of congruent elemental sound radiators, each differing only by successive delayed phase in order to steer the radiated sound beam pattern into a desired far field direction. The arrangement of the sensors/actuators on the beam structure is designed to instill such a unique set of elemental sound radiators. Numerical studies are performed to evaluate the effectiveness of the proposed sound steering concept into the far-field hemisphere. [Work supported by JSPS.]

2:15

2pSA2. Effects of material orientation on the radiation efficiency of baffled plates. J. Gregory McDaniel (Dept. of Aerosp. and Mech Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, jgm@bu.edu)

The orientation of an anisotropic material may have profound effects on acoustic radiation from the structure. The present work explores these effects for the case of a thin anisotropic plate in an infinite rigid baffle. An understanding of the acoustic radiation, in terms of the radiation efficiency, is begun by considering two effects. The first is that of wave steering, in which the direction of structural wave propagation is changed by the material orientation. The second is that of wave mixing, in which the the relative mixture of propagating and evanescent waves is changed by the material orientation. The effect on radiation follows from the fact that radiation from propagating and evanescent waves is known to be dramatically different. The relative importance of frequency and excitation is also explored in the context of these two effects. These results are expected to lead to improved designs of composite structures at no additional cost—simply by changing the orientation of the material. [Work supported by ONR.]

<center>2:30</center>

2pSA3. A method to measure the equivalent acoustic impedance of a scatterer. Seung-ha Lee (Ctr. for Noise and Vib. Control, Dept., of Mech. Eng., KAIST, Sci. Town, Daejeon, 305-701, Korea, seung-habird@kaist.ac.kr)

The procedures of simplifying scatterers' geometry and finding its equivalent acoustic impedance are proposed by using scattering holography and multiple multipole expansions. The scattered sound field depends on the incident field, the geometry, and the physical property of the scatterer, which can be measured by using tools such as a set of microphones and a 3-D scanner. Once the sound field over a surface is measured, the whole sound field of 3-D space can be reconstructed by acoustic holography. However, predicting near-field sound by holography is inefficient in practice, because huge numbers of measurement points are required to describe the complex sound field in the vicinity of a scatterer. With this regard, multiple multipole expansions are employed to reduce the number and the size of elements of a scatterer necessary to express the measured scattered field. Therefore it is equivalent to simplifying the geometry of a scatterer, and enables us to predict the sound pressure and velocity, as well as acoustic impedance, on the scatterers' surface. Using the proposed method, it is demonstrated that the equivalent acoustic impedance can be obtained. [Work supported by the BK21 project initiated by the Ministry of Education and Human Resources Development of Korea.]

<center>2:45</center>

2pSA4. Acoustic backscattering from fluid-filled submerged prolate spheroidal shells. Jeffrey E. Boisvert (NAVSEA Newport, Newport, RI 02841) and Sabih I. Hayek (Penn State Univ., University Park, PA 16802)

The equations of motion for nonaxisymmetric vibration of prolate spheroidal shells of constant thickness were derived using Hamilton's principle [S. I. Hayek and J. E. Boisvert, J. Acoust. Soc. Am. **114**, 2799–2811 (2003)]. The shell theory used in this derivation includes shear deformations and rotatory inertias. The shell displacements and rotations were expanded in infinite series of comparison functions. These include associated Legendre functions in terms of the prolate spheroidal angular coordinate and circular functions in the azimuthal angle coordinate. The fluid-filled shell is insonified by a plane wave with an arbitrary angle of incidence. The external and internal fluid loading impedances were computed using an eigenfunction expansion of prolate spheroidal wavefunctions. Far-field backscattered acoustic pressure spectra are presented as a function of the angle of incidence for several shell thickness-to-half-length ratios ranging from 0.005 to 0.1, and for various shape parameters, a, ranging from an elongated spheroidal shell ($a = 1.01$) to a spherical shell ($a \sim 100$). A comparison of the backscattering from fluid-filled and empty shells is presented at selected plane wave incident angles. [Work supported by the ONR/ASEE Summer Faculty Research Program and the NAVSEA Newport ILIR Program.]

<center>3:00</center>

2pSA5. Modeling a magnetorheological elastomer's transient stiffness parameters. Anne-Marie Lerner and Ken Cunefare (Georgia Tech., ken.cunefare@me.gatech.edu)

Semiactive vibration absorbers are tunable vibration absorbers whose natural frequencies can change controllably to minimize vibration better than their passive counterparts, which can be used to reduce occupant vibration in cars. The variable natural frequency is achieved by using a dynamically tunable spring. In order to use a material as a dynamically tunable spring, its transient parameters must be understood and controlled, so as not to lead to instability. This work characterizes the behavior of magnetorheological elastomers (MREs), which are a type of ferromagnetic-elastomer composites whose tensile/compressional stiffness increases as an applied magnetic field is increased. If the transition from one stiffness state to another is too slow, the MRE can have a detrimental effect on vibration control. Transient behavior of MREs with different concentrations of ferromagnetic material and different geometric shapes was recorded empirically, and the MRE stiffness change in response to a change in magnetic field was modeled as a function of ferromagnetic

content, MRE geometric shape, and the change in magnetic field strength. MREs in this configuration exhibited natural frequencies in the 30–65-Hz range, where MREs were able to change their natural frequency by up to 17%.

<center>3:15−3:30   Break</center>

<center>3:30</center>

2pSA6. Vibration absorption by an undamped beam element. Ilker Murat Koc, Adnan Akay (Dept. of Mech. Eng., Carnegie Mellon Univ., Pittsburgh, PA 15213), and Antonio Carcaterra (Univ. of Rome, La Sapienza, 00184, Rome, Italy)

Through two complementary approaches, using modal response and wave propagation, this presentation will show the conditions under which a decaying impulse response, or a nearly irreversible energy trapping, takes place in a linear conservative continuous system. The results show that the basic foundation of near-irreversibility or pseudo-damping rests upon the presence of singularity points in the modal density of the systems. To illustrate the concept of pseudo-damping in detail, a simple undamped beam is modified to introduce a singularity point in its modal density distribution. Simulations show that by a suitable application of a compressive axial force to an undamped beam on an elastic foundation, the impulse response of a beam shows the characteristics of a nearly irreversible energy trap and a decaying impulse response. Attaching such a modified beam to a single degree of freedom oscillator shows energy trapping by the beam producing pseudo damping. [Research carried out while AA served at NSF.]

<center>3:45</center>

2pSA7. Pseudo-damping in undamped plates and shells. Antonio Carcaterra (Dept. of Mech. and Aeronautics, Univ. of Rome, La Sapienza Via Eudossiana, 18, 00184, Rome, Italy) and Adnan Akay (Carnegie Mellon Univ., Pittsburgh, PA 15213)

Pseudo-damping is a counter-intuitive phenomenon observed in a special class of linear structures that exhibit an impulse response characterized by a decaying amplitude, even in the absence of any dissipation mechanism. The conserved energy remains within but designated parts of the system. Pseudo-damping develops when the natural frequency distribution of the system includes condensation points. The recently formulated theoretical foundation of this phenomenon, based on mathematical properties of special trigonometric series, makes it possible to describe a class of mechanical systems capable of displaying pseudo-damping characteristics. They include systems with discrete oscillators and one-dimensional continuous beamlike structures already reported by the authors in recent studies. This paper examines development of pseudo-damping phenomenon in two-dimensional structures, using plates and shells as examples, and shows how a preloaded plate on an elastic foundation can lead to pseudo-damping. Moreover, in the case of curved shell elements examined here, pseudo-damping can result due to the curvature of the structure, which naturally introduces condensation points in the modal density. [Research carried out while AA served at NSF.]

<center>4:00</center>

2pSA8. Fundamentals of apparent damping phenomena in linear conservative systems. Antonio Carcaterra (Dept. of Mech. and Aeronautics, Univ. of Rome, La Sapienza Via Eudossiana, 18, 00184, Rome, Italy) and Adnan Akay (Carnegie Mellon Univ., Pittsburgh, PA 15213)

This presentation addresses a class of irreversible phenomena that can develop in linear conservative systems and provides a theoretical foundation to explain the underlying principles. Recent studies have shown that energy can be introduced to a linear system with near irreversibility, or energy within a system can migrate to a subsystem nearly irreversibly, even in the absence of dissipation. Inspired by the properties of probability distribution functions, the general formulation developed here is based on particular properties of harmonic series, which form the common basis of

linear dynamic system models. The results demonstrate the existence of a special class of linear nondissipative dynamic systems that exhibit nearly irreversible energy exchange and possess a decaying impulse response. The formulation and its results also support the recent studies that observed near irreversibility and apparent damping in several dynamic systems and provide a common theoretical foundation for such behavior. [Research carried out while AA served at NSF.]

4:15

2pSA9. Using an enhanced constrained layer damping model to assess relative damping contributions. Zhengchao Xie and W. Steve Shepard , Jr. (Dept. of Mech. Eng., The Univ. of Alabama, Box 870276, Tuscaloosa, AL 35487, sshepard@eng.ua.edu)

Constrained layer damping (CLD) is an effective passive vibration control approach in which damping comes from the core layer's deformation. In the most general sense, this deformation will include shear, longitudinal extension, and transverse compression deformations. In this work, new analytical models and their single-layer finite-element (FE) implementations are used to examine the dependence of these types of damping on the frequency and the core layer's thickness for different boundary conditions. This information can be important in cases where a simplified model is needed with an acceptable error. The sandwich beam's response is examined by using various tools in order to determine the important aspects of the model, with particular emphasis on the relative importance of each of the three types of damping. Then, frequency responses found

using different single-layer finite elements are compared for different boundary conditions. By using both the analytical models and FE method results, guidance is provided on how to choose an appropriately simplified model or finite element for more practical analyses.

4:30

2pSA10. Effect of burst-modulated sinusoidal dither waveforms on the effectiveness of dither forces to eliminate friction-induced oscillations. Jeff W. Badertscher, Kenneth A. Cunefare, Al A. Ferri, and Michael M. Michaux (771 Ferst Dr., Rm. 205, Altanta, GA 30332)

This paper examines the effectiveness of high-frequency dither-cancellation techniques using burst-modulated signals. The classic single-degree-of-freedom (SDOF), mass on moving belt, model is analyzed using the method of averaging and numerical integration of system dynamics. Two different friction laws are used to model the contact point, both exhibiting a negative friction coefficient-velocity relationship. Recent results [M. Michaux and A. Ferri, Proceedings of IDETC05 Paper no. DETC2005-84491 (2005)] found that sinusoidal dither forces could stabilize or destabilize such a system, depending on the system and frictional characteristics as well as the dither signal amplitude and frequency. This paper extends this previous analysis to burst-modulated sinusoidal dither signals. Earlier experimental research has shown burst-modulated dither signals to be an effective method of suppressing automotive brake squeal. It is found that, for a given amplitude and frequency, burst signals are also capable of stabilizing the system. Individual effects of each burst parameter on system stability are established.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006   MOLOKAI ROOM, 1:00 TO 2:00 P.M.

Session 2pSCa

Speech Communication: First Language Acquisition (Poster Session)

Megha Sundara, Cochair
Univ. of Washington, Inst. for Learning and Brain Science, Box 357988, Seattle, WA 98195-7988

Kazuhiko Kakehi, Cochair
Chukyo Univ., Tokodachi 101, Toyota 470-0393, Japan

Contributed Papers

All posters will be on display and all contributors will be at their posters from 1:00 p.m. to 2:00 p.m. To allow for extended viewing time, posters will remain on display until 5:30 p.m.

2pSCa1. When is a phonological rule acquired? A look at children's production on syllable-final nasals in Taiwan Mandarin. Renee Hung (Dept. of English, Natl. Taiwan Normal Univ., 162 Hoping E. Rd., Sec. 1, Taipei, Taiwan 106, reneejh@hotmail.com) and Janice Fon (Natl. Taiwan Univ. Taipei, Taiwan 106)

Syllable-final nasals, /n/ and /N/, tend to be merged when they are after /i/ and ə in Taiwan Mandarin, but in different directions. For /i/, /n/ is merged with /N/, while for ə, /N/ is merged with /n/ (Chen, 1991; Tse, 1992; Hsu, 2005). This study thus aims to examine whether children in Taiwan merge syllable-final nasals in Mandarin the way adults do. Fifty Grade-2 and Grade-6 students would read monosyllabic words, containing 3 [vowels (/i/, ə, /a/)] ×4 (tones) =12 minimal pairs differing only in nasal endings. They would repeat the words twice, once in Zhuyinfuhao, a local phonetic system, and once in traditional Chinese characters. Preliminary results indicated that Grade-2 children treated the /eN/ words very similarly to adults, although the occurrences of nasal merge were less

frequent in the Zhuyinfuhao set than in the character set. For vowel /i/, however, merging of both directions could be found in both Mandarin characters and Zhuyinfuhao.

2pSCa2. Phonetic development of word-initial Korean stops in young Korean children. Minjung Kim and Carol Stoel-Gammon (Dept. of Speech & Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105-6246, minjungk@u.washington.edu)

A number of studies have investigated developmental patterns of stops in various languages through the instrumental analysis; however, very little is known about phonetic characteristics of Korean stops in speech of young Korean children. In Korean, stops are produced at three places of articulation (labial, alveolar, velar); at each place, there is a three-way contrast, yielding nine stops in word-initial position. All stops are voiceless and are differentiated by several acoustic parameters including VOT, f0 of the following vowel, and the amplitude difference of the first and

second harmonics (H1−H2) at the voice onset of the following vowel. The present study examines developmental aspects of word-initial Korean stops by measuring values of VOT, f0, and H1−H2 at the onset of the following vowel in words produced by Korean children aged 30, 36, 42, and 48 months. Findings will focus on the phonetic features used by Korean children to contrast the three-way stops and on comparisons between the phonetic patterns of stops produced by Korean children and Korean adults.

2pSCa3. Patterns of anticipatory lingual coarticulation in typically developing children.

Shawn Nissen, Kurtt Boucher, and Mathew Parson (Dept. of Commun. Disord., Brigham Young Univ., Provo, UT 84602)

Many studies have examined the patterns of speech coarticulation in adults and younger speakers; however previous findings have not established a consistent developmental perspective. There are contrasting views on whether children at early stages of development demonstrate more, less, or about the same amount of anticipatory coarticulation than adults. This study will focus on describing the patterns of anticipatory lingual coarticulation in the obstruent productions of three groups of children between 3 and 6 years of age ($N=42$) and one comparison group of adults ($N=14$). Frequency measures for the first and second formants were extracted from multiple tokens of a centralized and unstressed vowel (schwa) spoken prior to two different sets of productions. The first set of productions consisted of the target vowel followed by a series of real words containing an initial CV(C) syllable, while the second set consisted of nonword productions with a relatively constrained phonetic context. Formant values were transformed to a perceptually normalized scale (ERB) prior to statistical analysis. Analysis of variance was utilized to determine if the formant frequencies varied systematically as a function of age, gender, and phonetic context. [Work supported by research funding from Brigham Young University.]

2pSCa4. Developmental vowel acquisition data obtained by a band-filtering analysis method: Results from Hungarian children.

Krisztina Zajdó (Div. of Commun. Disord., Univ. of Wyoming, Dept. 3311, 1000 E. Univ. Ave., Laramie, WY 82071, kzajdo@uwyo.edu)

A recently developed band-filtering analysis method (van der Stelt *et al.*, 2003, 2005; Zajdó *et al.*, 2004) that minimizes the dependence of the results on fundamental frequency allows for a dependable identification of spectral measures in children's vowels. In the scripts, selection of segments with certain acoustic features (e.g., sound quality, $F0$) is followed by the calculation of spectral envelopes between 0–7 kHz, resulting in 40 filter values per spectrum. An intensity matrix is produced with data from each 40 filters. Data reduction is achieved through prinicipal component analysis. Results indicate that the first two principal components are related to the $F1$ and $F2$ of vowels, both in adults (Pols, 1977) and children (Zajdó *et al.*, 2005). Analyses of unlabeled vowel productions in Hungarian-speaking children at 2; 0, 3; 0 and 4; 0 years show that the first two eigenvectors account for $>50\%$ of the variability in all age groups. By using these eigenvectors, age-specific reference planes are created. Mapping phonemically labeled vowel measures onto the 4; 0-year-old's plane from each age group suggests that children produce decreasingly overlapping vowel categories as they age. With development, vowel categories are increasingly specified, covering more distinct areas of the acoustic vowel space. [Work supported by UW.]

2pSCa5. The acquisition of speech rhythm by bilingual Spanish- and English-speaking four- and five-year-old children.

Ferenc Bunta (Dept. of Commun. Sci., Temple Univ., Weiss Hall, 1701 North 13th St., Philadelphia, PA 19122, fbunta@temple.edu) and David Ingram (Arizona State Univ., Tempe, AZ 85287)

This study investigates the acquisition of speech rhythm by bilingual Spanish- and English-speaking children, comparing their performance to that of monolingual age-matched peers in both languages, and to monolingual and bilingual adults. Participants included younger children (between 3 years, 9 months and 4 years, 5 months, 15 days), older children (between 4 years, 6 months, 18 days and 5 years, 2 months), and adults (over 18 years). A sentence elicitation task of 26 sentences per language was used to obtain 64 samples that yielded over 15 000 duration measurements of vocalic and intervocalic intervals. The duration measurements were entered into the normalized vocalic and intervocalic pairwise variability index (PVI) measures based on Grabe and Low (2002). The results indicated that younger bilingual children displayed distinct speech rhythm patterns for their target languages and deviated from their monolingual English-speaking peers. Older bilingual children also produced significant separation of speech rhythm, and differences between older bilingual children and their monolingual peers speaking English were also found. Comparisons between younger and older bilingual children indicated an effect for the vocalic PVI measure, but not for the intervocalic PVI, providing partial support for age differences for bilingual children. Bilingual adults showed separation of their languages and performed similarly to their monolingual peers, overall.

2pSCa6. Comparative study of perception of Japanese dialectal accent in children.

Ayako Shirose (Waseda Univ., Mikajima 257915, Tokorozawa 3591129, Japan) and Kazuhiko Kakehi (Chukyo Univ., Tokodachi 101, Toyota 4700393, Japan)

This paper reports the acquisition course of perception of Japanese dialectal accent. We carried out an experiment of accent perception for native speakers of Japanese dialects, namely, Tokyo dialect (Standard Japanese) and Kagoshima dialect. Then we compared results between the two dialects. Experimental stimuli were real words with several types of synthesized accent patterns. Results revealed that the Kagoshima children and adults accepted both the native Kagoshima accent and the Tokyo accent, whereas the Tokyo children and adults accepted only the native Tokyo accent. These results suggest that the Kagoshima speakers had insensitivity to the perception of accent. In our previous studies, the Kagoshima children showed different characteristics of accent production from the Tokyo and Kyoto children: the Tokyo and Kyoto children acquired their native accent correctly, whereas the Kagoshima children failed to produce it. [Shirose *et al.*, J. Cognit. Sci. **2**(1), 33–44 (2001)]. The insensitivity of accent perception of both Kagoshima adults and children might be one factor contributing to results of accent production. These intriguing parallels of accent perception and production will be discussed inclusively.

2pSCa7. Infants' attention to repeated speech and musical patterns.

Gerald W. McRoberts (Haskins Labs., 300 George St., Ste. 900, New Haven, CT 06511, mcroberts@haskins.yale.edu) and Colin Dawson (Dept. of Psych., 1503 E Univ. Blvd., Univ. of Arizona, Tucson, AZ 85721)

Previous investigations have shown that infants older than 5 months listen longer to speech with repeated utterances than to speech without utterance repetition. The present studies investigated whether infants also attend preferentially to repeated musical patterns. Musical phrases of four and five notes and ending with a rest were composed using several pitch contours and rhythmic patterns. Experiments 1 and 2 compared trials in which phrases were immediately repeated (AABBCC), to trials in which the same phrases occurred without repetition (ABCDE). In the first experiment, no listening preference for either trial type was found for 6- or 9-month-old infants. In experiment 2, silence was added after each phrase to assist infants in parsing the phrases. Infants again failed to show a listening preference. In experiment 3, trials of speech, half with repeated utterances and half without repeated utterances (from previous experiments), were interspersed with trials of repeated and nonrepeated music. Nine-month-olds demonstrated a preference for the repeated trials of both speech and music over nonrepeated trials of the same type. The fact that a preference for musical repetition is only obtained through induction (from verbal repetition) suggests that repeated speech may hold a special attentional status for infants. [Work was supported by Grants NIH_R15/DC005947 and NIH_R01/DC00403.]

2pSCa8. Infants listening to songs: Exploring a processing trade-off between phonetic and pitch patterns. Gina C. Cardillo and Patricia K. Kuhl (Inst. for Learning and Brain Sci., Box 357988, Univ. of Washington, Seattle, WA 98195, ginacc@u.washington.edu)

This investigation explored 11-month-olds' responses to changes in the phonetic (lyric) and pitch (melody) content of songs. Sixty infants were tested using a between-subjects auditory preference paradigm. After being familiarized with a four-note song, infants were presented with both the familiar song and a new song that contained either a novel syllable or pitch order. In experiment 1, infants preferred listening to a novel melody when the phonetic content was redundant, but showed no preference when the identical melody change occurred in the presence of variegated phonetic content. In experiment 2, we tested the hypothesis that infants presented with songs containing variegated phonetic information would attend to changes in words but not melodies. Infants marginally preferred listening to novel songs when only the lyrics changed, but did not show a preference when only the melodies changed. Taken together, the results are consistent with *a priori* predictions, and suggest that even though infants can recognize melodic changes when they are presented in isolation, identical changes become more difficult to detect when more complex phonetic information is introduced. Thus, neural resources at this age may be more heavily committed to processing phonetic information than pitch information when both are present. [Work supported by the J.S.McDonnell Foundation, NIH.]

2pSCa9. Cognitive influences in infant speech perception. Barbara T. Conboy, Jessica Sommerville, Josie Randles, Caryn Deskines, and Patricia K. Kuhl (Inst. for Learning and Brain Sci., Univ. of Washington, UW Box 357988, Seattle, WA 98195-7988)

Speech perception during the first year reflects increasing attunement to native language phonetic features, but the mechanisms underlying this development are not well understood. Reductions in non-native phonetic discrimination have been linked to improvement in native phonetic discrimination and later vocabulary growth (Kuhl *et al.*, 2005), and performance on nonlinguistic tasks (Lalonde and Werker, 1995). The present study examined links between native and non-native voice onset time discrimination, receptive vocabulary (MacArthur-Bates CDI), and cognitive control abilities at 11 months. Infants ($n = 18$) completed a double-target conditioned head turn task and two nonlinguistic tasks requiring attentional control and resistance to irrelevant cues (means-end and detour-reaching object retrieval). Infants with CDI scores above the median showed higher native discrimination scores, $t(16) = 2.15$ 0.05, but no group differences for the nonnative contrast. Infants with scores above the median on either cognitive task showed worse discrimination of the nonnative contrast (means-end, $t(15) = 2.27 p 0.04$; detour-reaching, $t(15) = 3.49$, $p .01$), but no group differences for the native contrast. These results suggest that cognitive control plays a role in infants' ability to ignore acoustic cues that are irrelevant to their native languages phonemic categories. [Work supported by NICHD (Grant HD37954) and a UW NSF Science of Learning Center Grant (LIFE).]

2pSCa10. Distinguishing voices: Are infants' abilities affected by age or language experience? Megha Sundara and Patricia K. Kuhl (Inst. for Learning & Brain Sci., Univ. of Washington, Box 357988, Seattle, WA 98195-7988)

We know surprisingly little about infants' ability to encode voice information when listening to speech. Research with infants has focused on their ability to recognize their mother or fathers' voice and distinguish it from an unfamiliar voice. However, communication consists of interactions with unfamiliar speakers as well. Adults are well able to extract information about the speakers gender, size, age, dialect, etc., even from short samples of an unfamiliar persons voice. In experiment 1, we tested 4- and 12-month-old English-learning infants on their ability to distinguish an unfamiliar male and female voice when listening to Hindi monosyllabic words. Results indicate that the 12-, but not the 4-month-olds can distin-guish an unfamiliar male and female voice when listening to Hindi words. In experiment 2, we tested 4- and 12-month-old English-learning infants on their ability to distinguish an unfamiliar male and female voice when listening to English monosyllabic words. Now, both 12- and 4-month-olds are able to distinguish the voices. Together, these results indicate that infants' ability to distinguish voices is affected by age as well as their language experience. We are now testing adults to determine if the effects of language experience on the ability to distinguish voices persists even after development.

2pSCa11. The bilingual brain: A comparison of native and non-native speech perception in monolingual and bilingual infants. Patricia K. Kuhl, Sharon Coffey-Corina, Denise Padden, and Maritza Rivera-Gaxiola (Inst. for Learning & Brain Sci., Univ. of Washington, Seattle, WA 98195, pkkuhl@u.washington.edu)

Infants raised in monolingual families are equally good at native and non-native speech discrimination early in life. By 12 months, performance on native speech has significantly improved while non-native performance declines. We tested bilingual American infants at 7 and 11 months of age on native (/ta-pa/) and non-native (Mandarin affricate-fricative) contrasts used in the monolingual tests. Phonetic discrimination was assessed using behavioral (conditioned head-turn) and brain (event-related potential) measures. The MacArthur Communicative Development Inventory estimated infants developing language skills at 14, 18, 24, and 30 months of age. The monolingual data [Kuhl *et al.*, *Language Learning and Development* (2005)] demonstrated that at 7 months of age, infants native and non-native speech perception skills predicted their later language development, but differentially. Better *native* phonetic perception predicted more rapid language development between 14 and 30 months, whereas better *non-native* phonetic perception predicted slower language development over the same time period. Performance of bilingual infants on the same speech perception tasks, and their patterns of language growth between 14 and 30 months, will be compared to those of the monolingual infants, and a model that accounts for the results will be described. [Work supported by NSF, NIH, and the Human Frontiers Science Program.]

2pSCa12. Aussie, Aussie, Aussie, Oi-Oi-Oi: Infants love an Australian accent. Christine Kitamura (MARCS Auditory Labs., Univ. of Western Sydney, Locked Bag 1797, Penrith South DC NSW 1797, Australia, c.kitamura@uws.edu.au), Robin Panneton (Virginia Tech, Blacksburgm, VA 24061-0436), Anna Notley, and Catherine Best (Univ. of Western Sydney, Penrith South DC NSW 1797, Australia)

This study examines infant's attention to dialect differences following Nazzi *et al.*'s [J. Mem. Lang., **43**, 1–19 (2000)] finding that American 5-month-olds can discriminate British- and American-English. Using a serial preference task, 48 6-month-old Australian and American infants heard sentence sets spoken in Australian- and American-English. Results showed that at 6 months, American infants listen longer to Australian than American sentences, but Australian infants show no preference. By 8 months, American infants also show no preference. The developmental lag suggests Australian infants have more exposure to the American dialect (e.g., television programs) than American infants to the Australian dialect. Thus, it is predicted, with less linguistic experience, Australian 3-month-olds will show a dialect preference comparable to American 6-month-olds. Data from 14 3-month-olds support this hypothesis; moreover, they also listen longer to Australian than American sentences. Together the results imply that the ability to generalize across two dialects is a function of experience, and that, with age, infants filter out irrelevant phonetic information and cluster American and Australian dialects into the same group. Future directions of this study are discussed as are the reasons why Australian English is the dialect of preference for infants from both dialect environments.

2pSCa13. Word extraction in adult-directed and infant-directed speech in African-American-English-learning infants. Shiri Katz-Gershon and Jean Andruski (Dept. of Audiol. Speech Lang. Pathol., Wayne State Univ., 207 Rackham, 60 Farnsworth St., Detroit, MI 48202)

Although there is evidence that infant-directed speech (IDS) facilitates word extraction in Standard-American-English (SAE)-learning infants, this effect has not been tested in African-American-English (AAE)-learning infants. An important question that remains unaddressed is whether infants are better tuned to word boundary cues in their native dialect than a non-native dialect. Specifically, do infants extract words more efficiently from IDS in their native dialect than from IDS in a non-native dialect. This study compares the ability of eight-month-old AAE-learning infants to extract words from AAE and SAE IDS. The head-turn preference procedure was used to measure infants' memory for nonsense words embedded in sentences in AAE versus SAE. Infants were familiarized for 2 min with six sentences containing nonsense words and were then presented with both familiar and unfamiliar nonsense words carrying matching prosodic features. The difference in looking times between the familiar and unfamiliar nonsense words was used as a measure of word extraction. Results so far demonstrate an advantage of IDS over adult-directed speech in SAE. IDS in AAE is less advantageous than IDS in SAE. Acoustic analysis shows that the prosodic features of target nonsense words in SAE may make them more salient to infants than target words in the AAE sample.

2pSCa14. Exploring and learning the consequences of vocalizations in early infancy. Francisco Lacerda and Lisa Gustavsson (Dept. of Linguist. Stockholm Univ., SE-106 91 Stockholm, Sweden)

The ability to discover and learn about the consequences of own vocalizations is currently being studied in three groups of infants, 3–5-, 6–8-, and 9–11-month olds, exploring a situation that resembles an audiovisual memory game. The infants are tested individually, in a quiet and dimmed studio. They sit in front of a black computer monitor connected to an eye-tracking system. At randomly selected intervals, a circle appears in the center of the screen and expands and shrinks rapidly to a bright circle to signal that an observation trial is in progress. If the infant vocalizes during the trial, an animated actor emerges after about 200-ms delay at screen coordinates that are contingent on spectral characteristics of the initial part of the infant's vocalization. The actors brightness and time on display is controlled by the vocalizations overall intensity. By analyzing the changes in predictive looking behavior that probably occur during the session, we expect to gather valuable information on the infant's capacity to learn about consequences of their vocalizations at different ages. [Work supported by grants from the Bank of Sweden Tercentenary Foundation (MILLE, K2003-0867) and EU NEST program (CONTACT, 5010).]

2pSCa15. The acoustic characteristics of infant cries. Naoto Yamane and Yoko Shimura (The United Grad. School of Education, Tokyo Gakugei Univ., 4-1-1 Nukuikitamachi, Koganei, Tokyo, 184-8501, Japan, r052004s@u-gakugei.ac.jp)

The present study measured the acoustic characteristics of infant cries. Twenty-eight infants participated, ranging from 2 to 23 months of age. Crying was elicited when their mother moved to a distance, leaving them in the lap of unknown person. The cries were recorded with a microphone placed 1 m from the infant's mouth in an anechoic room. Analysis of the recorded cries revealed four patterns: cries were classified as either voiced or breathy, and as higher or lower than 7 kHz. Certain cry characteristics were later rated as more stress-causing, suggesting they could potentially elicit more abuse reactions. The recorded cries were presented to adults and breathy cries with a high-frequency component were labeled as emergency cries and as likely to cause high stress levels in parents and caregivers. In Japan, living spaces have recently begun changing such that environments are more echoic, perhaps exacerbating negative reactions to high-frequency components of infant cries.

2pSCa16. Discovering place and manner features—What can be learned from acoustic and articulatory data? Ying Lin (Dept. of Linguist, Univ. of Arizona, Tucson, AZ 85721) and Jeff Mielke (Univ. of Ottawa, Ottawa)

Theories of phonological acquisition often assume that features are innate and universal, and establishing distinctive features requires the availability of minimal pairs. In this study, we explore an inductive approach to feature discovery without resorting to minimal pairs. Our basic assumption is that learners have access to a wide variety of phonetic information, and so feature induction can proceed through recursive clustering of the data. To test our hypothesis, we used as our data acoustic and articulatory measurements made simultaneously from a set of CVC syllables. The articulatory data are obtained with ultrasound imaging, and the vocal tract is characterized by a vector of cross distances from the tongue body to a reconstructed palate. The raw articulatory data are then mapped to a lower dimensional space using principle components, on which clustering is carried out. Experiments using acoustic and articulatory data suggest that these two sources of information lead to the discovery of place and manner features, respectively, thus bringing support to an alternative theory of feature learning.

Session 2pSCb

Speech Communication and Psychological and Physiological Acoustics: Second Language Acquisition I

Jared C. Bernstein, Cochair

Ordinate Corporation, 800 El Camino Real, Ste. 400, Menlo Park, CA 94025

Reiko Akahane-Yamada, Cochair

ATR, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Chair's Introduction—2:15

Invited Papers

2:20

2pSCb1. Second-language speech perception: The modification of automatic selective perceptual routines. Winifred Strange (Ph.D. Program in Speech and Hearing Sci., City Univ. of New York Grad. School, 365 Fifth Ave., New York, NY 10016, strangepin@aol.com)

It is well established that late second language (L2) learners show significant difficulties in categorizing L2 consonants and vowels that do not occur in the native language (L1), or that occur, but differ in their phonetic realization in L1 and L2. These perceptual problems underlie (some) L2 production difficulties (accented L2 speech); lead to L2 comprehension difficulties, especially in nonoptimal listening environments; and may persist even after years of immersion in the L2 environment. However, it has been shown that intensive laboratory training (listening with feedback) can lead to improved perception of difficult non-native contrasts that: (a) generalizes to novel speakers and stimulus materials, (b) persists over months, and (c) produces improvements in production of trained segments. In the model presented here, changes with experience in patterns of phonetic perception are conceived of as reflecting the modification of language-specific (L1) selective-perceptual routines used to extract phonologically relevant information efficiently from continuous speech input. Theoretical and methodological issues will be discussed regarding (a) the conceptualization of (changes in) the structure of phonetic/phonological categories, (b) empirical assessment of phonetic perceptual learning, (c) operational definitions of mastery of L2 phonological categorization, and (d) implications for improving phonetic perceptual training techniques. [Work supported by NIDCD.]

2:40

2pSCb2. STRAIGHT as a research tool for L2 study: How to manipulate segmental and supra-segmental features. Hideki Kawahara (Faculty of Systems Eng., Wakayama Univ., 930 Sakaedani, Wakayama 640-8510, Japan) and Reiko Akahane-Yamada (ATR, Sorakugun, Kyoto 619-0288, Japan)

A morphing algorithm based on STRAIGHT speech manipulation system [Kawahara *et. al.*, ICASSP (2003)] has a potential importance in L2 study, since it generates stimulus continuum with quality close to natural speech. The morphing procedure yields an equally spaced continuum in an abstract parameter domain, which is spanned by fundamental frequency, STRAIGHT's time-frequency representation (a variant of smoothed spectrogram), the periodicity map, and the time-frequency coordinate. Instead of manipulating all these parameters at once, it is useful to manipulate each parameter independently. This individual parameter manipulation can also be generalized to coupled manipulation by introducing new parameters defined using a set of parameters. This extension enables us to manipulate segmental and supra-segmental features instead of manipulating abstract physical parameters. For example, coupling energy envelope, fundamental frequency, and temporal axis stretching factor yield access to a prosodic feature that is basically supra-segmental. Manipulation of spectral landmark positions yields access to a phonological feature that is basically segmental. A preliminary test results using the right-light continuum sets generated using a STRAIGHT-based method and the Klatt synthesizer will also be presented.

3:00

2pSCb3. Structural representation of the pronunciation and its application to computer-aided language learning. Nobuaki Minematsu (Grad. School of Frontier Sci., The Univ. of Tokyo, 5-1-5, Kashiwanowa, Kashiwa-shi, Chiba, 277-8562, Japan)

The conventional speech applications have been developed by looking at sound substances of speech and this strategy is derived from acoustic phonetics. However, the speech representation of this framework, e.g., spectrogram, is inevitably with nonlinguistic factors such as speakers and microphones. Then, all the developed systems come to select users, where the systems may work fine with a major part of the users but still work poorly with the others. The author believes that this situation should be avoided for educational applications because users don't know whether they are outliers or not. Recently, the author proposed another framework of developing speech applications by looking at only sound contrasts and this strategy is derived from physically and mathematically interpreting structural phonology. On this new framework, nonlinguistic factors such as speakers and microphones are mathematically

3137   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3137**

2p WED. PM

removed from speech as pitch information can be removed by smoothing the spectrogram. This talk shows what is possible by the new framework, which completely discards the absolute acoustic properties such as formants and spectral envelopes. In other words, in the new framework, an utterance is modeled as organized pattern of sounds although it has been modeled just as a linear string of sounds in the conventional framework.

3:20

2pSCb4. Using speech recognition for just-in-time language learning. Maxine Eskenazi, Antoine Raux, and Elizabeth Harris (Carnegie Mellon Univ., 4619 Newell Simon Hall, 5000 Forbes Ave., Pittsburgh, PA 15213)

While many language learning systems rely on the student to be a dedicated learner, the opportunity to teach language is present outside the classic teaching situation, in all everyday oral and text-related activities. Teaching language in real situations, where the new knowledge is needed and used immediately, has great potential for robust learning: items are retained over a long period of time, make learning other items easier, and are generalized to other situations. Giving immediate and effective error detection and corrective feedback is a technical challenge for language modeling, spoken dialogue structure, and targeted speech synthesis. We will discuss just-in-time language learning and then present one type of just-in-time learning, lexical entrainment, where the automatic system detects an oral error and then respeaks the utterance that contained the error in corrected form, emphasizing the corrected portion. To test the effectiveness of this approach, the Lets Go spoken dialogue system that furnishes bus information to Pittsburghers was modified to predict errors from a variety of non-native speakers of English. A study of the effectiveness of lexical entrainment for a group of non-native users will be described. [This work is sponsored by NSF Grants 0208835 and IIS0096139.]

3:40

2pSCb5. Theory and method in automatic evaluation of spoken language proficiency. Jared Bernstein (Ordinate Corp., 800 El Camino Real, Menlo Park, CA 94025)

Several kinds of theory (linguistic, psycholinguistic, and sociolinguistic) have been used to justify different methods of language proficiency assessment or to explain patterns of performance observed in proficiency tests. After reviewing some traditional methods of testing spoken language performance (e.g., structured interviews, listening comprehension, word identification in noise, and elicited imitation), the assumptions underlying these methods are compared. When the practicality, reliability, and validity of assessment methods are evaluated, differences in the associated theories become clearer. A low perplexity testing method is presented that scores a set of context-free speech-interactive tasks using speech recognition and analysis techniques to measure aspects of the linguistic and phonetic content of spoken responses. Data from several experiments comparing interview testing with low-perplexity testing on populations of second language speakers (of English, Spanish, and Dutch) suggest that nearly all the information that expert judges extract about levels of communicative competence can be accounted for in terms of automaticity of spoken language processing. Furthermore, measurement techniques that are feasible in low-perplexity testing provide much greater precision in the spoken language proficiency reporting scale than was previously feasible, opening up a range of new applications.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006

HONOLULU ROOM, 1:00 TO 5:00 P.M.

Session 2pSP

Signal Processing in Acoustics: Spatial Speech, Non-Destructive, Time Reversal

Elizabeth A. McLaughlin, Cochair
Naval Undersea Warfare Center, 1176 Howell St., Newport, RI 02841

Yoshifumi Chisaki, Cochair
Kumamoto Univ., Faculty of Engineering, 2-39-1 Kurokami, Kumamoto 860-8555, Japan

Contributed Papers

1:00

2pSP1. Evaluation of spatial impression of sound in horizontal and depth directions for the purpose of developing an advanced multichannel live sound reproduction system. Hiroyuki Okubo, Toshiyuki Nishiguchi, Reiko Okumura, and Kimio Hamasaki (Human & Information Sci., NHK Sci. & Tech. Res. Labs., 1-10-11 Kinuta Setagaya, Tokyo 157-8510 Japan)

A prototype of an advanced multichannel live sound reproduction system was developed. In this system, multiple loudspeakers are arranged in the horizontal and depth directions in front of listeners to reproduce a sense of sound depth and width [K. Hamasaki *et al.*, Proc. 119th AES Convention, (2005) p. 6559]. Using this prototype, a number of subjective

tests were conducted to evaluate the listeners sense of sound depth and width. The results revealed that while the direct sound source was perceived near the predicted horizontal location, the sense of depth varied largely from subject to subject. Another listening test was conducted in which reflections from the upper-front of the listeners were reproduced. Simulated reflections of a virtual space were divided into 11 directional room impulse responses (RIRs). The RIRs were convoluted with individual sound sources (male vocals), and subjects compared their spatial impressions of the sound sources reproduced with and without such reflections. The results of this listening test revealed that the feeling of sound depth was one of the factors of spatial impressions and it increased when the reflections from the upper-frontal direction were reproduced.

1:15

2pSP2. Precise position measurement of objects using a coded acoustic signal in an indoor environment. Kuramitsu Nishihara (Grad. School of Sci. and Technol., Chiba Univ., 1-33 Yayoi-cho Inage-ku, Chiba, 263-8522 Japan), Tadashi Yamaguchi, and Hiroyuki Hachiya (Chiba Univ., Chiba, Japan)

Acoustic sensing in the air might serve as an effective technique for acquiring information about an object, such as its distance, shape, and surface configuration. Nevertheless, it is difficult to identify acoustic waves clearly because many mirror reflections or environmental noises exist in the air. In addition, acoustic measurement requires a longer time than measurements using lasers or electromagnetic waves because sound propagates much more slowly than light. For this study, we attempted highly precise position measurement using an M-sequence signal in an indoor environment. The degrees of M-sequence were 7–14; the transmitting frequency was 25 kHz. A signal reflected from various angles was received by moving the receiving unit, which contained three microphones. Measurement results with the M-sequence signal showed that the signal-to-noise ratio (S/N) improved 20–40 dB compared with a burst wave. Furthermore, the distance resolution improved when the M-sequence degree was changed. The distance and the angle of multiple objects installed indoors or in walls were detectable as a result of construction of a two-dimensional plane based on received signals.

1:30

2pSP3. Self-similar analysis of the room impulse response. Yong Tang, Hideo Shibayam, Wakako Tanaka, and Tooru Itakura (Dept. of Commun. Eng., Shibaura Inst. of Technol., 3-7-5 Toyosu Koutou-ku, Tokyo, 135-8548 Japan, m603101@sic.shibaura-it.ac.jp)

Important information of the room acoustic characteristics is contained in the room impulse response. We analyze the time-space correlation of the impulse response by the self-similarity method. The room impulse response is divided into the frame of the length of the fixed time. Attractors are composed by affine transformation of each frame. The attractors that make the Hausdorff distance become the shortest between the standard frame and attractors are decided. The impulse response is reconstructed by using these attractors. In the self-similar analysis, the frame length is decided from the period of the Hausdorff distance between the initial sound and the reflection sound. We show results for the relation between the room acoustic characteristics and the attractors calculated from the self-similar analysis.

1:45

2pSP4. Application of dynamic synapse neural networks on gunshot recognition. Alireza A. Dibazar, Hyung Ook Park, Walter M. Yamada, and Theodore W. Berger (Dept. of Biomed. Eng., Univ. of Southern California, 1042 Downey Way, DRB140, Los Angeles, CA 90089-1111, dibazar@usc.edu)

Identification of acoustic signals in noisy environments remains one of the most difficult of signal processing problems, and is a major obstacle to the high degree of accuracy and speed needed to identify suspicious sounds in high-security, high-safety environments. We have previously developed an acoustic recognition capability using a novel, biologically based Dynamic Synapse Neural Network (DSNN) technology. The DSNN-based technology has been demonstrated to classify target sounds with a high degree of accuracy, even in high noise conditions. In this paper we focus on extending the acoustic recognition capability of DSNNs to the problem of gunshot recognition. In order to recognize and localize the event, an array of four microphones is used for sound input. For localization purpose, time-delay estimation algorithm (TDE) is employed for triangulation. We have developed stand-alone, portable, and cost-efficient hardware with which both recognition and localization can be performed. In field-testing, the system classifies and localizes over 90% of the trained-for sounds. Performance for firecracker, starter pistol, 9-mm, and 44-caliber, explosion/firing sounds was also tested.

2:00

2pSP5. Effects of reproduction methods and audio coding on word intelligibility with competing talkers. Kenji Ozawa and Junichi Nakagai (Interdisciplinary Grad. Sch. of Med. and Eng., Univ. of Yamanashi, 4-3-11 Takeda, Kofu, 400-8511 Japan, ozawa@yamanashi.ac.jp)

Effects of reproduction methods and high-efficiency audio coding on word intelligibility were examined in a tele-collaboration system. Monaural, stereophonic, and binaural methods were studied. Transform domain weighted interleave vector quantization (TwinVQ) [Iwakami *et al.*, Proc. ICASSP'95, pp. 937–940 (1995)] was used as the audio coding method. Intelligibility tests using word lists, in which word familiarity was controlled, were conducted under six conditions combining three reproduction methods with two conditions, i.e., with or without audio coding. In the experiments, three talkers pronounced different words simultaneously; a listener was then asked to selectively hear the word of a designated talker. Analyses of variance of the experimental results showed that intelligibility for the binaural method was higher than for the other two reproduction methods in both conditions with or without audio coding. Moreover, the benefit of the binaural method was demonstrated because intelligibility for the method with audio coding was higher than that for the stereophonic method without coding, whereas the amount of audio data in the former method was much smaller than in the latter, indicating that the binaural method with high-efficiency audio coding is useful for a tele-collaboration system. [Work supported by JSPS: KAKENHI (17560336) and RIEC, Tohoku Univ. (H18/A10).]

2:15

2pSP6. Environmental adaptation of hidden Markov models in an automatic sounds recognition system used for surveillance applications. Asma Rabaoui, Zied Lachiri, and Noureddine Ellouze (ENIT (Ecole Nationale des Ingenieurs de Tunis) Rue Ibn Mandhour, Residence Sana, Bloc Yasmine, App10, ARIANA 2080 Tunisie)

Generally, the main goal of an automatic sound recognition system is to analyze in real time the sound environment of a habitat and to detect the abnormal sounds that could indicate a distress situation. In a surveillance application, the most encountered problem is the background noise that can be captured with the sounds to be identified. Thus, we propose to use a multistyle training system based on HMMs: one recognizer is trained on a database including different levels of background noises and is used as a universal recognizer for every environment. In order to enhance the system robustness, we will explore different adaptation algorithms that reduce the environmental variability. The algorithms closely examined are maximum likelihood linear regression (MLLR), maximum *a posteriori* (MAP), and the MAP/MLLR algorithm that combines MAP and MLLR. Experimental evaluation illustrates a recognition improvement over the baseline system (i.e., none adapted system) results. By using 156 adaptation scenes, the error rate of the adapted recognizer improves over the baseline system by more than 30% for rather all the considered algorithms, even under important noise degradation conditions.

2:30

2pSP7. Passive acoustic fiber Bragg grating sensor measurements of Green function in an instrumented part. Karl A. Fisher, Sean K. Lehman, and David H. Chambers (Lawrence Livermore Natl. Lab., L-333, 7000 East Ave., Livermore, CA 94550)

The development of fiber Bragg grating (FBG) sensors for acoustic measurements permits the instrumentation of manufactured composite parts for the purposes of nondestructive evaluation (NDE) and lifetime monitoring. The parts can be designed from conception with an array of FBG acoustic receivers integrated into its internal structure for the purpose of multistatic data collection for NDE and monitoring of internal changes. The FBG array is accessed via a port on the outer surface and requires no additional power demands on the operation of the part. Data collection is accomplished by activating sources on the surface of the part and measuring the fields along the FBG array. We are particularly interested in Green function (GF) estimation of a FBG instrumented composite structure. By

passively measuring diffuse acoustic fields throughout the volume, we use a technique developed by Lobkis and Weaver [J. Acoust. Soc. Am. **110**(6), 3011–3017 (2001)] to estimate the GF within the part and monitor it over time. We describe FBG sensor array operation, review the theory of diffuse field GF estimation, and present examples in heterogeneous multilayered structures. [This work was performed under the auspices of the U.S. DOE by University of California Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.]

2:45

2pSP8. Super-resolution algorithms for ultrasonic nondestructive evaluation imagery. Grace Clark (Univ. of California Lawrence Livermore Natl. Lab., 7000 East Ave., L-130, Livermore, CA 94550, clark9@llnl.gov) and Jessie Jackson (Univ. of California Lawrence Livermore Natl. Lab., Livermore, CA 94550)

The spatio-temporal range resolution of ultrasonic measurements is often severely limited by the inherent band-limited spectral transfer function of ultrasonic transducers. This is manifested as ringing in the time domain. Super-resolution algorithms and a new MATLAB-based software system are presented for improving spatial resolution in one-dimensional (1-D) pulse-echo signals (A-scans). Two-dimensional (2-D) images (B-scans) and 3-D volumes are formed from suites of processed 1-D signals. The algorithms solve two inverse problems that are ill-posed and ill-conditioned. Regularization techniques and prior constraints are used to ensure useful results in practice. Given a noisy reflection signal (A-scan) $u(t)$ and an associated reference signal $x(t)$, the algorithms first solve the system identification problem to produce an optimal least-squares band-limited estimate $\hat{h}(t)$ of the impulse response of the material under test. Next, an algorithm based upon the method of alternating orthogonal projections produces a spectrum−extrapolated version $\hat{h}_e(t)$ of the impulse response estimate. This estimate, when used in place of the raw reflection signal $u(t)$, enhances the spatio−temporal resolution by mitigating the transducer distortion. The efficacy of the algorithms is demonstrated on real data from controlled experiments. Three different practical applications are presented, showing improved location of material interface boundaries. [Work supported by Grant UCRL−ABS−222494.]

3:00−3:15 Break

3:15

2pSP9. Nondestructive technique to detect a crack in a concrete block by precise estimation of the surface wave. Takeshi Murakami, Toyota Fujioka, Yoshifumi Nagata, and Masato Abe (Dept. of Comput. and Information Sci., Faculty of Eng., Iwate Univ., Ueda 4-3-5, Morioka City, Iwate, Japan)

This paper describes a method to estimate a crack's position in a concrete block using several accelerometers. An array of accelerometers is attached to the concrete block and an impact is made on the concrete surface with a small hammer. If a crack exists, a reflection wave is generated at the crack position. Conventional methods to estimate the reflection point may seem to be useful for detection. Since the concrete block is elastic, however, it has three wave propagation modes: the surface wave mode, the primary wave mode, and the secondary wave mode. We cannot estimate the position using conventional methods because the necessary primary wave mode is weaker than the surface wave mode. To estimate the crack position precisely, we have already proposed two methods to eliminate the surface wave and side wall reflections. However, since the elimination is insufficient, they sometimes yield a peak at a position where there is no crack. Therefore, this paper proposes a method to estimate the surface wave more precisely to suppress such a peak. Some experiments were carried out, yielding better results.

3:30

2pSP10. Applying broadband beam-based decomposition of the time reversal operator (BBDORT) to midfrequency active sonar systems. David M. Fromm (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5350)

This paper reports on progress developing broadband beam-based decomposition of the time-reversal operator (BBDORT) for midfrequency hull-mounted active sonar systems. BBDORT combined with backpropagation has been previously demonstrated to localize a mid-water-column target in shallow water using a vertical source-receiver array [Fromm and Gaumond, J. Acous. Soc. Am. **118**, 2039]. However, the array spanned the water column, and choosing an optimal beam set (to scan for targets in a short period of time) can make as much as 30-dB target improvement over nonoptimal beam sets. Current research efforts will address: source-receiver array configurations and tactical frequencies of Navy operational interest, evaluating performance in different environments, and determining efficient algorithms to optimize beam sets. [Work supported by the Office of Naval Research.]

3:45

2pSP11. Properties of focused time reversal operator in random media. Jean-Luc Robert (Philips Res. North America, 345 Scarborough Rd., Briarcliff Manor, NY 10510, jean-luc.robert@philips.com) and Mathias Fink (Universite Paris 7, 75005 Paris, France)

FDORT (focused DORT) is an alternative of the DORT method using focused transmits instead of single-element transmits. In the case of well-resolved point scatterers, the FDORT eigenvectors may be understood as invariants of the time reversal operator. However, if the medium is a collection of many subresolution scatterers (speckle), a different interpretation is more helpful. It is shown that, in speckle, the FDORT matrix may be interpreted statistically as a spatial correlation matrix, whose coefficients are the correlation between signals received by pairs of elements. The matrix is fully described by the Van Cittert Zernike theorem. The first eigenvector still has focusing properties, and the eigenvalues can be related to the so-called focusing criterion. Applications to aberration estimation in medical ultrasound are presented.

4:00

2pSP12. Spatio-temporal invariants of the time-reversal operator. Jean-Luc Robert (Philips Res. North America, 345 Scarborough Rd., Briarcliff Manor, NY 10510, jean-luc.robert@philips.com) and Mathias Fink (Universite Paris7/ESPCI, 75005 Paris, France)

The DORT method gives the spatial invariants of the monochromatic time-reversal operator, at a given temporal frequency. It allows the monochromatic Green functions of point scatterers to be extracted from a data set. However, in acoustics, broadband signals are usually used and thus, it is often interesting to have the invariants in the time domain, or to know the evolution of the monochromatic Green function with frequency. It is not always easy, especially in complex media, because the sequence of the DORT eigenvectors (the invariants) changes as a function of the frequency. Thus, an invariant corresponding to a given scatterer may be the first at low frequencies, but the second at higher frequencies. Therefore, it is difficult to reconstruct a spatio-temporal Green function from monochromatic invariants. We introduce here the spatio-temporal time-reversal operator as a linear operator described by a fourth-order tensor, and deduce a method to obtain its spatio-temporal invariants. The method reduces to a more straightforward matrix decomposition, and thus can be easily implemented. Properties of the invariants are described, and applications to focusing in complex media are presented.

4:15

2pSP13. Semiautomatic inverse filtering of vowels with differing degrees of nasality. Michelle Tooher and Christer Gobl (Phonet. and Speech Lab., School of Linguistic Speech and Commun. Sci., Trinity College Dublin, Dublin 2, Ireland, mtooher@tcd.ie)

Most approaches to inverse filtering are either manual or automatic. Manual methods, while providing high-quality fine-grained analysis, require user expertise and manpower to analyze even small amounts of data. Automatic methods allow large-scale analysis, but lack the precision required to capture non-neutral voice dynamics as present in, for example, expressive speech. This paper reports on development of a semiautomatic inverse-filtering toolkit, focusing on an adaptive nasal/non-nasal vowel inverse filter. For vowels with differing degrees of nasality (from not nasalized to heavily nasalized), each frame is analyzed using machine learning techniques to assess the degree of nasality. Based on this assessment, the inverse filter is adapted according to acoustic theory, to best analyze each frame of the data, employing either all-pole or pole-zero analysis where necessary. User intervention is possible, in both specifying analysis frames and fine-tuning of pole and zero number and location. Training data for machine learning are selected on the basis of auditory and acoustic analysis. This semiautomatic approach allows analysis of large-scale corpora while retaining the robustness and fine-grained analysis of manual methods. Results from analysing both non-nasalized and nasalized vowels from a corpus of Irish are presented. [This work is funded by the IRCSET Embark Initiative.]

4:30

2pSP14. Ambiguity analysis of matched-field source localization under spatially correlated noise field. Wen Xu (Teledyne RD Instruments, 14020 Stowe Dr., Poway, CA 92064)

Matched-field processing (MFP) concerns estimation of source location by exploiting full-wave modeling of acoustic waveguide propagation. The ambiguity output in MFP relies on the correlation between the signal field at the true source position and the signal field at each scanning source position. This correlation often shows a multimodal structure due to the nonlinear parameter dependence: in addition to a mainlobe around the true source position, there are unpredictable prominent high sidelobes elsewhere. For a well-conditioned problem, in the absence of noise the peak output is guaranteed to occur at the true source position. In the presence of noise, we will probably have some peak outputs around the sidelobes, introducing a large localization error. Therefore, output ambiguity structure is a very important factor in the development of any matched-field algorithm operated in low-SNR scenarios. To analyze the ambiguity behavior, a quantitative approach for error analysis has previously been developed in the context of the maximum likelihood estimate (MLE) with spatially white noise. In this talk, generalization to including spatially correlated noise field introduced by discrete interferences and/or surface distribution sources is discussed.

4:45

2pSP15. Multiuser communications using passive time reversal. Heechun Song, William S. Hodgkiss, William A. Kuperman, Tuncay Akal (MPL/SIO, UCSD, La Jolla, CA 92093-0238), and Mark Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

A recent paper [Song *et al.*, IEEE J. Ocean. Eng. **31**, 165–173 (2006)] demonstrated multipleinput/multiple-output (MIMO) multi user communications in shallow water using active time reversal where the time reversal array (i.e., base station) sent different messages to multiple users simultaneously over a common bandwidth channel. Passive time reversal essentially is equivalent to active time reversal with the communications link being in the opposite direction. This paper describes MIMO passive time reversal communications, which enables multiple users to send information simultaneously to the time reversal array. Experimental results at 3.5 kHz with a 1-kHz bandwidth demonstrate that as many as six users can transmit information over a 4-km range in 120-m-deep shallow water using QPSK modulation, achieving an aggregate data rate of 6 kbits/s. Moreover, the same data rate has been achieved at 20 km range by three users using 16-QAM modulation.

WEDNESDAY AFTERNOON, 29 NOVEMBER 2006   MAUI ROOM, 1:00 TO 5:30 P.M.

Session 2pUW

Underwater Acoustics and Acoustical Oceanography: Sediment Acoustic Processes: Quantifying the Effects of Sediment Properties on Dispersion, Attenuation, Reflection, Scattering, and Buried Target Detection II

Eric I. Thorsos, Cochair

Univ. of Washington, Applied Phys. Lab., 1013 NE 40th St., Seattle, WA 98105-6606

Masao Kimura, Cochair

Tokai Univ., Dept. of Geo-Environmental Technology, 3-20-1 Shimizu-Orido, Shizuoka 424-8610, Japan

Contributed Papers

1:00

2pUW1. Concerning the sediment volume contribution to scattering in SAX04 (Sediment Acoustics Experiment 2004) sediment: Sound speed and bulk density fluctuations. Kevin Briggs (Seafloor Sci. Branch, Naval Res. Lab., Stennis Space Ctr., MS 39529-5004)

Sound speed and bulk density profiles measured from SAX04 sediment cores are used to estimate correlation lengths of fluctuations in the sound speed and the bulk density. Sound speed is measured from 5.9-cm-diameter cores at 1-cm intervals and bulk density is measured at 2-cm intervals from the same cores. Previous estimates of these parameters from experimental data involved fewer realizations (cores), limited sampling areas, and samples without mud flasers within the sand sediment. Values of correlation length may vary according to location within the SAX04 site. Correlation length values estimated with Burg's algorithm from sediment sound-speed fluctuations averaged 2.24 cm, from sediment bulk density fluctuations averaged 2.53 cm. Ranges of values of correlation length estimated from bulk density fluctuations from locations inside the experiment site are used to present a range of volume scattering strength predictions, given the other parameters of sediment sound speed, sound attenuation, density ratio variance, and ratio of sound speed to density fluctuations used by the perturbation approximation to the composite roughness model for each location. At the SAX04 site, interface scattering

dominated over volume scattering, despite the presence of mud inclusions with strong sound speed and bulk density contrasts. [Work sponsored by ONR.]

1:15

2pUW2. The effects of layering on reflection coefficient and transmission measurements from a sand/water interface at the Sediment Acoustics Experiment 2004 (SAX04). Marcia Isakson (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

During the Sediment Acoustics Experiment 2004 (SAX04), transmission measurements from 4.5 to 50 kHz were taken on an eight-transducer array buried at an angle of 15 deg using a source mounted on a remotely operated vehicle (ROV). Simultaneously, the reflection coefficient was measured using a three-receiver vertical line array (VLA). Reflection coefficient measurements at angles from 10 to 70 deg were obtained. Environmental characterization of the area has revealed layering effects such as a surficial mud layer, mud inclusions, and depth-dependent sediment properties such as porosity. The consequence of this layering on the variability of the amplitude of reflection coefficient measurements and the time of arrival of transmission measurements will be investigated. [Work supported by ONR, Ocean Acoustics.]

1:30

2pUW3. Measurement and modeling of elliptical particle motion in the seabed. John C. Osler, David M. F. Chapman, Paul C. Hines, Jeffrey Scrutton (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, john.osler@drdc-rddc.gc.ca), and Anthony P. Lyons (The Penn State Univ., State College, PA 16804)

As part of the SAX04 experiment conducted off the coast of Florida in the Gulf of Mexico, four vector sensors containing three-axis accelerometers and pressure sensors were buried in the seabed. These served as the receivers to measure sediment sound-speed dispersion using a variety of techniques. One technique involved the generation of spherical waves from a point source in water and transmitted into the seabed in the frequency band 800 to 3000 Hz. In this geometry, the contribution of inhomogeneous waves to the field results in elliptical—rather than longitudinal—particle motion. The orientation of the elliptical orbit varies with the source-receiver geometry, frequency, and the sediment sound speed, thereby allowing measurements of sediment sound-speed dispersion. However, the orientation of the ellipse major axis is not always aligned with the direction of wave propagation suggested by Snell's law. Measurements were made at several angles approaching the nominal critical angle in order to address the trade-off between increased sensitivity of the measurement versus greater departure from Snell's law. Measurements and modeling of the elliptical particle motion are compared to ensure that the sediment sound-speed estimates account for any bias created by the inhomogeneous waves. [Work partially supported by ONR.]

1:45

2pUW4. Characterization and modeling of synthetic aperture sonar image statistics. Anthony P. Lyons, Shawn F. Johnson, Douglas A. Abraham (The Penn State Univ., Appl. Res. Lab., State College, PA 16804, apl2@psu.edu), and Kevin L. Williams (Univ. of Washington, Seattle, WA 98105)

Interest in the application of synthetic aperture sonar (SAS) to seafloor mapping and object detection has been increasing in recent years. Because images of the seafloor obtained with SAS systems can sometimes exhibit large pixel-to-pixel fluctuations, robust seafloor segmentation or target detection still requires a thorough understanding of the SAS image statistics and their relation to seafloor parameters. As part of the SAX04 sediment acoustics experiment, acoustic data suitable for SAS processing were collected using a bottom-mounted rail system off the coast of Florida in the Gulf of Mexico in October, 2004. This unique data set has been used to explore several important aspects of SAS image statistics including their dependence on system resolution and on the correlation structure of SAS reverberation due to ripples that were present at the study site. Models of

the reverberation envelope probability distribution function (pdf) based on the K-distribution have been found to well describe the reverberation environment for systems with lower resolutions than possible with broadband SAS systems. Characterization and modeling of SAS image statistics based on the K-distribution and using ground truth collected during SAX04 will be presented. [Work performed under ONR Grants N00014-04-1-0013 and N00014-06-1-0245.]

2:00

2pUW5. Modeling the effect of multipath on high-resolution synthetic aperture sonar image statistics. Shawn F. Johnson, Anthony P. Lyons, Douglas A. Abraham (Appl. Res. Lab., The Penn State Univ., State College, PA 16804, sfjohnson@psu.edu), and Kevin Williams (Univ. of Washington, Seattle, WA 98105)

Recent analysis of Synthetic Aperture Sonar data collected during the Office of Naval Research sponsored SAX04 experiment in the Gulf of Mexico during October 2004 has shown that the processed image has varying amplitude statistics as a function of range. Since synthetic-aperture sonar maintains a constant range-independent resolution cell size, unlike traditional sonar systems, one would expect that the statistics of the image would be based solely on the seafloor properties. The K-distribution shape parameter was estimated as a function of range and increases at discrete ranges corresponding to the onset of multi-path propagation (low shape parameter values represent a heavy tailed distribution). At these points of increase in shape parameter, scattered returns arising from additional propagation paths arrive with the result that two or more parts of the seafloor are contributing to the same sonar resolution cell, which leads to a larger K-distribution shape parameter than would be expected when only the direct path contributes. Model results using the sonar system geometry and interface scattering theory will be used to illustrate this effect. [Work supported by ONR Codes 321 and 333. First author supported by a National Defense Science and Engineering Graduate Fellowship.]

2:15

2pUW6. Synthetic aperture sonar imaging of buried spheres and cylinders. Steven G. Kargl and Kevin L. Williams (Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105)

During SAX04, synthetic aperture sonar (SAS) was used to detect buried spheres where the nominal grazing angle of incidence from the SAS to the point above a sphere was well below the sediment's critical grazing angle. SAS images from other measurements below the critical angle on buried spheres and finite cylinders in the NSWC-PC test pond facility have also produced target detections. Numerical simulations will be discussed that accurately predict the observed signal to noise (SNR). Simulations include estimates of reverberation from the rough seafloor, and the subcritical penetration through the seafloor, scattering from a target, and propagation back to the SAS. Required material parameters were obtained from environmental measurements including profiles of the small-scale surface roughness and the superimposed ripple structure. These simulations, the sonar equation results, and the model/measurement comparisons over the frequency range of 10–50 kHz further support sediment ripple structure as the dominant mechanism for subcritical penetration in this range. [Work supported by ONR.]

2:30

2pUW7. Measurement of scattering wave from a buried object using an underwater vehicle. Koichiro Tanaka, Taku Suto, and Takahiro Fukai (Ishikawajima-Harima Heavy Industries Co., Ltd., 1, Shin-Nakahara-Cho, Isogo-Ku, Yokohama 235-8501, Japan, kouichirou_tanaka@ihi.co.jp)

Measurements of a scattering wave from a buried object were carried out at a site in Tokyo Bay. The buried object was set below the bottom with about 10-m water depth. A crane suspended a subbottom profiler to insonify the object. In these measurements, the medium frequency (less than 10 kHz) was used to penetrate near the critical angle. Also, when the object shape is a cylinder and so on, the scattering wave appears strongly

in the specific direction, which is with the shape of the object and the physical relationship of the source and the object. The direction where this scattering wave appears is strongly different from the source position, and it is difficult to catch in the mono-static position. Especially, because the level of the scattering wave becomes low in the case of the scattering from the object under the seafloor, it is favorable if the scattering wave can be measured in the direction with this strong level. Therefore, to catch this strong scattering wave, an underwater vehicle was used. A hydrophone array was attached at the head of the underwater vehicle to accumulate over a series of pings. This report deals with the way and results of these measurements.

2:45

2pUW8. Buried target detection with a synthetic aperture sonar. John Piper (NSWC-PC, 110 Vernon Ave., Panama City, FL 32407)

During SAX04 the NSWC-PC dual-frequency synthetic aperture sonar collected data from a buried target field. Detections of a buried fluid-filled sphere from above critical angle averaged 34 dB and from below critical angle averaged 24 dB with a ripple field with predominant 42-cm spacing. These results are compared to simulated PCSWAT results.

3:00−3:15   Break

3:15

2pUW9. Using a dual-band synthetic aperture sonar for imaging various seafloor compositions. John S. Stroud, Daniel Brown, Jr., Daniel Cook, Jose Fernandez (NSWC-PC Code HS11, 110 Vernon Ave., Panama City, FL 32407), Kerry Commander (Office of Naval Res., Arlington, VA 22203), David Kolesar, and Tom Montgomery (Appl. Res. Lab, State College, PA 16804)

The Small Synthetic Aperture Minehunter (SSAM) is a synthetic aperture sonar mounted aboard a 12.75-in. REMUS600 unmanned underwater vehicle transmitting both low (10–50 kHz) and high (105–135 kHz) frequency bands. The sonar and the vehicle are highly integrated, which allows precise pulse timing, attitude correction, and object geolocation. Operating simultaneously in both bands produces two independent images of each scene. The upper band is capable of 2.5- by 2.5-cm resolution that provides high-fidelity images of proud targets. The lower band, processed to 7.5- by 7.5-cm resolution, adds clutter reduction as well as capabilities against buried targets. The dual bands of this sonar have proved to be a robust solution for minehunting in the littorals. In the past 18 months, SSAM has been deployed nine times. In these tests, the system has operated against bottom types including mud, sand, rippled sand, layered mud and sand, posidonia, and coral. Imagery from sea tests showing all of these bottom types will be presented. [The authors would like to acknowledge support from the ONR, Code 32.]

3:30

2pUW10. Performance of a buried object scanning sonar against targets buried in sand at the SAX-04 (Sediment Acoustics Experiment 2004) site. Steven Schock (Dept. of Ocean Eng., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431, schock@oe.fau.edu)

The performance of synthetic aperture processing for imaging buried targets in the band of 3 to 20 kHz is quantified by measuring the SNR of 3-D acoustic images of targets buried in sand at the SAX-04 site. A buried object scanning sonar consisting of an omnidirectional projector and a 40-channel hydrophone line array is towed within 10 m of the seabed at various horizontal offsets to buried targets to measure the SNR of target echoes as a function of grazing angle. Detections of the targets occurring at grazing angles smaller than the critical angle suggest that buried targets can be detected using the evanescent boundary wave thereby improving the range of the sonar for detecting shallow-buried targets. Acoustic images of buried targets also show that backscattering from sand ripples, sediment variability, and sediment layer interfaces limit the performance of the sonar for detecting and imaging buried objects.

3:45

2pUW11. The design of a boundary element package for treating acoustic scattering from buried objects. Ralf Burgschweiger, Martin Ochmann (Univ. of Appl. Sci., Dept. of Mathematics, Phys., Chemistry, TFH Berlin Luxemburger Strasse 10, D-13353 Berlin, Germany), and Bodo Nolte (Federal Armed Forces Underwater Acoust. and Marine Geophys. Res. Inst. (FWG), D-24148 Kiel, Germany)

The target strength of objects placed in water and partially buried in the sediment are calculated by means of a boundary element solver. For this reason, a boundary element code with suitable pre- and postprocessor is developed. With the help of the preprocessor, the scattering objects within the transition range between fluid and sediment can be modeled. The preprocessor can manage large and complex meshes (more than 100 000 elements) and is able to import NASTRAN and ANSYS mesh-files. It automatically generates closed geometric meshes. These meshes are used as database for the BEM calculations. For reducing computing time, the boundary element solver runs on a high number of personal computers in parallel equipped with Windows- or Linux-operating systems. These calculator hosts will receive their task from the preprocessor using TCP/IP-based interprocess communication. The parallel calculations can be performed as a service in the background and remotely administrated by the postprocessor. The postprocessor polls the results from the calculator hosts, delivers them to the appropriate projects, and provides visualization and analysis of the results. The efficiency of such a parallel processing boundary element package will be demonstrated for the acoustic scattering at simple test objects consisting of a high number of surface elements.

4:00

2pUW12. Subcritical grazing angle detection of targets buried under a rippled sand interface. Joseph L. Lopes, Carrie L. Dowdy, Raymond Lim (Naval Surface Warfare Ctr., Code HS11, 110 Vernon Ave., Panama City, FL 32444), Dajun Tang, Kevin L. Williams, and Eric I. Thorsos (Univ. of Washington, Seattle, WA 98105-6698)

Controlled acoustic measurements were conducted to investigate subcritical detection of targets buried in sand having a rippled sediment-water interface. The measurements were performed in a freshwater test-pond with a 1.5-m layer of sand on the bottom. Bottom ripple was formed by scraping the sand with a machined rake. Broadband (10 to 50 kHz) transducers were placed onto an elevated rail that translated horizontally permitting data to be processed using synthetic aperture sonar (SAS) techniques. Targets included silicone oil-filled spheres and flat-endcapped, solid aluminum cylinders. Data were acquired for various ripple wavelengths, heights, and orientations. For each bottom configuration, the seabed roughness over the buried target was measured to verify the ripple parameters and to estimate the small-scale roughness spectrum. This information was used in models predicting backscatter signal levels from the target and bottom. Results for spherical targets have generally been good. However, when cylindrical targets were inspected, model agreement was variable for reasons not fully understood. Characteristics of the target return are found to depend sensitively on ripple height and wavelength, and bottom parameters. Results are discussed with an eye towards understanding previously unexplained field test detections of buried cylindrical targets. [Work supported by ONR and SERDP.]

4:15

2pUW13. Scattering of evanescent waves incident on targets in a simulated sediment. Curtis F. Osterhoudt and Philip L. Marston (Dept. of Phys. and Astron., Washington State Univ., Pullman, WA 99163)

When a sound beam in water is incident on a sediment at a sufficiently small grazing angle, the resulting wave field in the sediment is complicated, even for the case of flat, fluidlike sediments. The wave field in the sediment for a sound beam from a simple transducer has an evanescent component and diffractive components. This situation has been experimentally simulated by using a combination of previously described immiscible liquids [Osterhoudt *et al.*, J. Acoust. Soc. Am. **117**, 2483 (2005)]. Measured wave-field features are similar to those seen in a wave-number-

integration-based synthesis (using OASES) for a related problem. Targets are placed in the evanescent wave field to simulate scattering by targets buried in marine sediments. Features of the observed backscattering are relevant to detecting and clarifying objects in sediments. Measurements include the resonant excitation of the organ-pipe mode of open water-filled cylinders previously studied with free-field illumination [Osterhoudt and Marston, J. Acoust. Soc. Am. **113**, 2334 (2003)]. The degradation of the backscattering with increasing target depth and with target tilt were measured and modeled. Simple approximate models of these properties are facilitated by the localized coupling processes for the excitation of (and radiation by) organ-pipe modes. [Work supported by ONR.]

4:30

2pUW14. Observations of subcritical acoustic penetration of a sandy seafloor. Darrell R. Jackson, Brian T. Hefner, Kevin L. Williams, and Eric I. Thorsos (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, drj@apl.washington.edu)

During the sediment acoustic experiments, SAX99 and SAX04, buried hydrophone arrays were used in conjunction with sources mounted on a movable tower to study subcritical acoustic penetration. The seafloor at the SAX99 site displayed prominent, regular sand ripples of 48-cm wavelength. These caused downward diffraction of sound incident below the critical angle, resulting in elevated sound-pressure fields in the sediment. These measurements were made over the frequency range 11–50 kHz and were in substantial agreement with a first-order perturbation model. For SAX04, the frequency range was extended down to 2 kHz, but the site had weaker, more disorganized ripples. As expected, this resulted in weak acoustic penetration, with the pressure field primarily due to the evanescent wave. To further test the penetration model, artificial ripples of wavelength 50 and 75 cm were created by divers. These ripples increased sound penetration levels by 5–10 dB when the incidence azimuthal direction was normal to the ripple crests. The diffraction angles of the penetrating sound were found to be consistent with the predictions of the perturbation model. [Work supported by ONR.]

4:45

2pUW15. Backscattering from sediments: Comparison of scattering strengths from a simple sand and from a sand/mud environment. Kevin Williams, Eric Thorsos, Dajun Tang, and Darrell Jackson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, williams@apl.washington.edu)

Two sediment acoustics experiments were carried out within 1 km of each other off the coast of Florida, the first in 1999 and the second in 2004. Though close in space the sediment was very different due to the effects of Hurricane Ivan in 2004. Hurricane Ivan resulted in the deposition of mud patches on top of the medium sand. Subsequent weather events deposited sand on top of these patches. The backscattering at each site will be presented for frequencies from 20 to 1000 kHz and grazing angles from 20 to 45 deg. At angles below the critical angle (30 deg) and frequencies below 150 kHz the backscattering measured in the two experiments is consistent given the difference in the sediment roughness spectra.

Above the critical angle, however, the backscattering measured in the 2004 experiment is several dB higher than in the 1999 experiment and increases immediately above the critical angle instead of decreasing. This difference is attributed to an increase in sediment volume scattering in 2004. From 150 to 1000 kHz the data from both years indicate a change in scattering mechanism below the critical angle to a mechanism differing from sediment interface roughness scattering. [Work supported by ONR.]

5:00

2pUW16. Measurements of interface roughness and examination of near bottom interface properties. Dajun Tang, Brian Hefner, Kevin Williams, and Eric Thorsos (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

A series of measurements was made of sediment properties during SAX04 over a 40-day period using IMP2, a sediment conductivity system; a stereo camera system made fine-scale interface roughness measurements toward the end of this period. During this period, several weather events (including Hurricane Ivan and a subsequent tropical storm) altered the sediment properties. The large, nonsymmetric ripples resulting from the hurricane decayed to a more symmetric form over time, while mud suspended in the water column was found first on top of the sand and then trapped just beneath the interface. Comparison of the interface roughness power spectra from IMP2 and the stereo camera shows good agreement at overlapping spatial scales and the stereo measurements extend the IMP2 estimation of the power spectra to a much higher spatial frequency. Complementary cores taken by a team from the Naval Research Laboratory at the same sites where IMP2 was deployed confirmed the depth and location of the imbedded mud found by IMP2. The implications of these environmental measurements to acoustic backscattering will be discussed. [Work supported by the ONR.]

5:15

2pUW17. Laboratory study of high-frequency scattering from water-saturated granular sediments. Anatoliy Ivakin (APL-UW, 1013 NE 40th St., Seattle WA 98105), Jean-Pierre Sessarego, and Regine Guillermin (CNRS/LMA, 13402 Marseille cedex 20, France)

Sound backscattering and reflection from water-saturated granular sediments at frequencies from 200 kHz to 7 MHz were studied in controlled laboratory conditions. Two kinds of well-sorted sandy sediments, fine and coarse sands, and two kinds of glass beads with corresponding sizes were chosen for the study. The two types of sand had narrow log-normal size distributions of particles with the mean diameters 0.25 and 1.5 mm for fine and coarse sand, respectively. The sediments were degassed and their surface was flattened carefully. In these conditions, the grain-scattering mechanism can be considered as a dominating factor controlling incoherent component of the field scattered from the sediment. Frequency dependencies for the backscattering strength at various grazing angles and the reflection coefficient at normal incidence were measured. The effects related to the sediment grain size are analyzed and their possible applications to remote sensing of marine sediments are discussed. [Work supported by ONR and CNRS.]

Session 3aAAa

Architectural Acoustics, Structural Acoustics and Vibration, Noise and Engineering Acoustics: Recent Developments in Acoustical Materials and Structures

Brandon D. Tinianov, Cochair

Quiet Solution, 1250 Elko Dr., Sunnyvale, CA 94089-2213

Kimihiro Sakagami, Cochair

Kobe Univ., Environmental Acoustics Lab., Faculty of Engineering, Rokkodai, Nada, Kobe 657-8501, Japan

Kirill V. Horoshenkov, Cochair

Univ. of Bradford, School of Engineering Design and Technology, Bradford, BD7 1DP, U.K.

Chair's Introduction—7:30
Invited Papers

7:35

3aAAa1. Characterizing viscoelastic and anisotropic porous materials. Laurens Boeckx, Poonam Khurana, Gerrit Vermeir, Walter Lauriks (Laboratorium voor Akoestiek en Thermische Fysica, Celestijnenlaan 200D, BE-3001, Leuven, Belgium, laurens.boeckx@fys.kuleuven.be), and Wim Desmet (Katholieke Universiteit Leuven, BE-3001, Belgium)

A wide range of commercial applications (building acoustics, food industry, automotive industry) can be found for porous materials. Accurate material characterization and modeling is vital for the use of these materials in multilayered systems due to increasing demands in acoustic comfort, noise legislation, and quality control during production. There is, however, a distressing lack in raw, accurate material data and measurement methods concerning the characterization of these materials. Measuring and modeling methods for the characterization of these materials will be presented. The experimental technique for the determination of the elastic properties is based upon the excitation of waveguides in porous materials. The other parameters of the Biot-Allard model are predicted using ultrasound. Acoustical properties can be predicted by using the measured structural material properties and incorporating them into a transfer-matrix-based multilayered model.

7:55

3aAAa2. Linking microstructure and acoustic properties of open-cell foams. Camille Perrot, Raymond Panneton (GAUS, Dept. of Mech. Eng., Universite de Sherbrooke, QC, Canada, J1K 2R1, camille.perrot@usherbrooke.ca), and Xavier Olny (ENTPE DGCB URA CRNS, Audin 69518, Vaulx en Velin, France)

A research program has been initiated in 2002 in order to link microstructure of high porosity open-cell foams to their acoustic properties. This paper is intended to highlight the main results of this study. The general objective of the research program is the determination of the acoustical macro-behavior from the physics at the local scale. A real rigid-framed porous media is studied. To this end, one needs first to determine the local geometry of the media, and second to solve over this geometry the partial differential equations that govern dissipation phenomena by thermal and viscous effects. The first step has been overcome by the technique of computed microtomography. This leads to experimental identification of the parameters of an idealized periodic unit-cell. The second step, solving harmonic heat and viscous fluid equations, is performed using Brownian motion and finite element simulations, respectively. Then, macroscopic behavior is obtained by spatial averaging of the local and frequency-dependent thermal and velocity fields. Results are presented in terms of two dynamic characteristic functions (viscous and thermal permeabilities) compared to impedance tube measurements. This computational methodology may be seen as a first step to optimize the microstructure of foams from a bottom-up approach for better sound proofing.

8:15

3aAAa3. Interlaboratory experiments on the characterization of the acoustical and related nonacoustical properties of porous media. Kirill Horoshenkov, Amir Khan (Univ. of Bradford, Bradford, UK), Frank Sgard, Francois Xavier, BecoLuc Jaouen, Amlie Renault, Nesrine Amirouche (Ecole Nationale des Travaux, Lyon, France), Jorn Hubelt (Gesellschaft fr Akustikforschung Dresden mbH (AFD), Germany), Francesco Pompoli, Nicola Prodi, Paolo Bonfiglio (Universita di Ferrara, Ferrara, Italy), Walter Lauriks, Laurens Boeckx (Katholieke Universiteit Leuven, Belgium), Giulio Pispola, Francesco Asdrubali (Univ. of Perugia, Perugia, Italy), Noureddine Atalla, Celse Amdin (Univ. of Sherbrooke, Sherbrooke, Canada), K. Mark Gmerek, and Adam Weston (The Boeing Co., Chicago, IL)

A series of reproducibility experiments on the characterization of acoustical parameters of selected samples of porous media is carried out on a range of porous samples in several independent laboratories in Europe and North America. The data on the characteristic acoustic impedance and complex propagation constant are presented in this work. In addition, the assessment of the related geometrical parameters required for modeling the acoustic performance of porous media, namely the steady-state flow resistivity, porosity, tortuosity, viscous and thermal characteristic lengths, and thermal permeability, is carried out. Detailed procedures related to sample preparation, and installation are discussed together with data on the material property variation observed between individual material samples and laboratories.

3aAAa4. Sound absorption characteristics of a honeycomb-backed microperforated panel (MPP) absorber. Kimihiro Sakagami, Kosuke Nakajima, Masayuki Morimoto (Environ. Acoust. Lab., Faculty of Eng., Kobe Univ., Rokkodai, Nada, Kobe, 657-8501, Japan), Motoki Yairi, and Atsuo Minemura (Kajima Corp., Chofu, Tokyo, 182-0036, Japan)

Microperforated panels (MPPs) are typically made of a thin metal or plastic panel and are often unsuitable for an interior finish because thin limp panels do not have enough strength. In particular, an interior finish of room walls requires appropriate strength. In order to solve this problem, a honeycomb structure is attached behind MPPs to stiffen the construction. Thus, it is possible to stiffen an MPP without increasing its thickness, which is important to keep MPPs at their best absorption performance. Furthermore, a honeycomb can increase MPPs' absorption coefficient in a similar way as a porous layer backed by a honeycomb. In this study, an experiment was performed to gain insight into the acoustical effect of a honeycomb structure behind MPPs and a simple theoretical model to interpret the experimental effects is presented. The experimental results show that the honeycomb affects the absorption characteristics of MPPs: the absorption peak increases and shifts to lower frequencies. This effect becomes more significant as the thickness of the honeycomb increases. The results from the theoretical model show the same tendency. This is attributed to the fact that the honeycomb makes a similar condition to local reaction in the back cavity.

3aAAa5. Development of new sound insulators with perforated board and honeycomb layer systems. Masahiro Toyoda and Daiji Takahashi (U. and E. Eng., Grad. School of Eng., Kyoto Univ., Katsura C1-4-392, Nishigyo-ku, Kyoto 615-8540, Japan, ae.toyoda@archi.kyoto-u.ac.jp)

Two newly developed types of sound insulators are introduced in this study. The first type is derived from an analytical model of a vibrating surface with an impedance facing. The model is investigated theoretically; the results indicate the possibility of reducing radiation from the vibrating surface by giving appropriate impedance. To realize this effect, a model using a perforated board with a subdivided air cavity is proposed. It is shown theoretically and experimentally that this insulator can achieve radiation reduction at an arbitrary frequency. The second type is proposed from the viewpoint of the subdivision strategy. It presents the possibility of improving insulation by restricting the air-particle motion at the interface between the vibrating surface and air. This method for noise control has an attractive simplicity and considerable practical benefit. The attenuation mechanism is discussed theoretically and experimentally. Results of this investigation show that the effect of this insulator is characterized by reduction of acoustic radiation at low frequencies near and below the critical frequency of coincidence. Wide use of these two insulators in many fields is anticipated.

3aAAa6. The effects of damped panels on sound transmission loss through lightweight framed construction. Brandon Tinianov (Quiet Solution, 1250 Elko Dr., Sunnyvale, CA 94089)

The incorporation of damping into the design of materials has been long understood as an effective method for improving the transmission loss of undamped, lightweight structures. Classic examples of its application include marine vessels, aircraft, and ground transportation such as trains, trucks, and buses. However, damping has rarely been incorporated into the materials used for the lightweight construction techniques found in modern buildings. This paper will review the basic principles of designing improved damping into construction materials using the well-known RKU model. The presentation will also review the mechanisms of energy transfer through lightweight building systems and present the improvements in acoustical performance due to damped materials. Typical gains are from 3–18 points in single number ratings (STC, Rw) and greater than 20 dB in some 1/3-octave bands. Extensive laboratory transmission loss data will be presented.

Contributed Papers

3aAAa7. Resilient ceiling construction in residential buildings. Anthony Nash (Charles M. Salter Assoc., 130 Sutter St., San Francisco, CA 94104, anthony.nash@cmsalter.com)

In multifamily buildings, the floor/ceiling assembly between living units needs to be sound rated. In North America, a common light-frame floor construction involves wood or cold-formed metal joists spaced at 400 mm. The typical practice is to attach a gypsum board ceiling to the underside of the floor framing using narrow strips of "Z"-shaped thin metal called "resilient channels." If the resilient channels or the ceiling are installed incorrectly, the low-frequency sound insulation of the assembly can be seriously degraded. Based on our experience with resilient ceiling construction, there are two significant factors that have not yet been documented in an acoustical laboratory. One factor is to maintain a small separation around the entire perimeter of the ceiling where it intersects vertical wall surfaces (this void is later sealed airtight with a flexible caulking material). A second is the particular design of the Z-shaped resilient channel. The paper will discuss findings from two acoustical test series of floor/ceiling assemblies in wood-frame buildings. In both cases, the assembly was tested before and after modifications to the perimeter of the existing resilient ceiling. Each of the series also included tests of a complete ceiling reconstruction using an improved resilient channel design.

3aAAa8. The effect of damping materials on heavy-weight floor impact in apartment buildings. Jin Yong Jeon, Seung Yup Yoo, and Pyoung Jik Lee (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr)

Floating floors, in which resilient isolators are inserted between the structural slab and the upper layer of the floor, are generally used because of their effectiveness in controlling structure-borne and airborne noise. However, these isolators amplify low-frequency noises below 100 Hz that are generally produced by heavyweight impacts. Viscoelastic damping materials are widely used to reduce noise in settings such as vehicles, ships, and machinery; however, there has been no report of their use in apartment building structures for reducing floor impact sounds. The addition of damping can be embodied using viscoelastic liquid such as silicone gasoline; energy is absorbed with superior characteristics being a solid matter of viscoelastic materials. In this study, optimized thickness of the damping layer was investigated for the efficient reduction of low-frequency noise generated by heavy-weight floor impact. Effects of temperature variations on the characteristics of storage modulus and loss factor of the material were also investigated.

10:15

3aAAa9. Sound absorbing material made of polyester nonwovens.
Shin-ichi Koga   (Fukoku Co., Ltd., 254 Nakayama, Mitsuhashi-machi, Yanagawa, 832-0811 Japan), Kazutoshi Fujimoto, and Ken Anai   (Kyushu Univ., Hakozaki, Higashi-ku, Fukuoka 812-8581 Japan)

Polyester nonwovens (PW) comprise fine fibers, similarly to glass wool (GW). Therefore, PW is expected to have high sound-absorption capabilities. In recent years, application of PW, an entirely recyclable material, as an absorbent material has attracted much attention from the viewpoints of resource recycling and waste reduction. In some cases, manufactured PW has been used as an interior material in an actual acoustic laboratory. Reportedly, however, PW has slightly inferior sound absorption capabilities compared to GW. For wider use as a sound-absorbing material, PW must have better sound absorption. In this study, sound absorptions of PW were measured for various fiber compositions. Specifications that engender high sound absorption capabilities were examined. Results revealed PW specifications that provide similar sound absorption as GW with similar thickness and density. The PW presented in this paper has both high sound absorption and good environmental properties. Therefore, it is useful as a sound absorbing material not only for architectural use, but also in various other fields. Finally, two examples are introduced of an actual anechoic chamber that was finished using PW.

10:30

3aAAa10. Sound-absorbing material made of recycled polyester nonwovens using scrapped pieces.   Kazutoshi Fujimoto, Ken Anai, Tatsurari Nakano   (Faculty of Human-Environment Studies, Kyushu Univ., 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581 Japan), and Shin-ichi Koga   (Fukoku Co., Ltd., Mitsuhashi-machi, Yanagawa, 832-0811 Japan)

The authors developed polyester nonwovens (PW) that have nearly equivalent sound absorption to glass wool with similar thickness and density. Unfortunately, scraps amounting to 20% of crude polyester are generated during PW manufacturing and cutting. They must be discarded as waste materials. Therefore, the authors contrived an original recycling technique to use such scrap pieces. Sound absorption characteristics of recycled materials manufactured using this method (RPW) are inferred to differ from those of PW because of their different fiber compositions. In this study, sound absorption of RPW is measured and the following points are examined experimentally: the relationship between manufacturing processes and sound absorption of RPW; sound absorption of RPW when two kinds of scrap pieces are mixed; and effects of mixing crude materials on RPW sound absorption. Results indicate that RPW has sufficient sound absorption for practical use. Comprising scrapped pieces that have been discarded as waste, RPW addresses the social importance of good resources usage and waste reduction; it also offers advantages of lower cost than PW. This RPW material is anticipated for use in various fields as a sound-absorbing material.

10:45

3aAAa11. Propagation characteristics of airborne ultrasonic wave in porous materials.   Kenichi Aoki and Tomoo Kamakura   (The Univ. of Electro-Commun., 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan, aoki@ee.uec.ac.jp)

The attenuation coefficient and propagation speed of airborne ultrasound are measured for highly porous open-cell polyurethane foams and fibers at frequencies of 1 kHz to 1.7 MHz. A theoretical model is proposed to explain the frequency responses of the transmission loss and speed of an air coupled wave physically in porous materials. The model is derived from Biot's flow resistance and density [M. A. Biot, J. Acoust. Soc. Am. **28**, 179–191 (1956)], Lambert's bulk modulus of fluids in pores [R. F. Lambert, J. Acoust. Soc. Am. **72**, 879–887 (1982)], and Zwikker and Kosten's concept for compliance of the side holes with entrance resistance [*Sound Absorbing Materials* (Elsevier, Amsterdam, 1949)]. Using measured data of static flow resistance to determine the mean pore size and the proposed model, theoretical prediction is made for the transmission losses and sound speeds in porous materials. Good agreement of theory and results of experiments in the whole frequency range supports the usefulness of the present model. Additionally, the model provides some findings for the extra attenuation coefficient of a slow wave in cemented glass bead specimens and sandstone for high-frequency ranges [P. B. Nagy, J. Acoust. Soc. Am. **93**, 3224–3234 (1993)].

11:00

3aAAa12. Comparative results of field impact isolation testing of numerous resilient underlayment systems for hard-surfaced flooring.
Marlund Hale   (Advance Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065, mehale@aol.com)

There is a strong ongoing interest in replacing carpet and pad flooring in existing multi-family residences and during apartment conversions to condominiums with hard surface flooring. Installing such flooring often results in a significant decrease in the impact isolation performance of the floor/ceiling assembly. Some municipalities and/or homeowners associations have enacted regulations requiring that floor/ceiling assemblies meet specific minimum laboratory and/or field acoustical performance standards. This paper presents published vendor data and field performance test results for two condominiums in a multi-family residential complex for which both the city and the homeowners association have such performance requirements. A series of field tests was undertaken to assist the homeowners in selecting qualifying underlayment systems for their preferred hard surface flooring replacement of their existing quality carpet and pad flooring. The test flooring systems were each installed by a flooring contractor. All aspects of these field tests were maintained as identical as possible, with the only variable being the different commercial resilient underlayment products. Published vendor data of certified laboratory IIC test results are compared with the FIIC field test results.

11:15

3aAAa13. Reflectivity of absorptive surfaces.   Jonathan Rathsam and Peter D'Antonio   (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

The random incidence absorption coefficient is a very coarse metric. In many small and large room applications, a directional absorption coefficient more accurately describes potential problems from first-bounce specular reflections. An impedance tube is useful for evaluating normal incidence absorption, but a more general approach to determining a material's specular reflectivity or first-bound attenuation is needed. The AES-4id-2001 information document describes such a method. This study compares the "first-bounce specular attenuation" of 1 ft (H) $\times$ 3 ft (W) and 4 ft (H) $\times$ 8 ft (W) samples of 2-in.-thick, 6 pcf fiberglass absorbers (with and without fabric covering) with that of a rigid reflector, using a 2D boundary method goniometer. The boundary wave's arrival is calculated and identified in the impulse response measurements. The use of small and large samples allows evaluation of both the fabric covering and boundary wave effects on the specular attenuation, for normal and 45-deg incidence. Time and frequency domain results are presented.

11:30–11:50
Panel Discussion

3a THU. AM

Session 3aAAb

Architectural Acoustics: Acoustics of Large Indoor and Outdoor Performance Spaces II

Daniel R. Raichel, Cochair
2727 Moore Ln., Fort Collins, CO 80526

Toshiki Hanyu, Cochair
Nihon Univ., Narashinodai, Funabashi, Chiba 274-8501, Japan

Contributed Papers

8:00

3aAAb1. Acoustics of ancient Greek and Roman theaters in use today. Anders Christian Gade and Konstantinos Angelakis (Acoust. Technol., Tech. Univ. of Denmark, Bldg. 352, DK 2800 Lyngby, Denmark, acg@oersted.dtu.dk)

In the Mediteranan area a large number of open, ancient Greek and Roman theatres are still today facing a busy schedule of performances including both classical and contemporary works of dance, drama, concerts, and opera. During the EU funded ''Erato'' project and a subsequent master thesis project, extensive measurement data have been collected from three well-preserved theatres, which represent three different stages of enclosing the audience in an open-air environment: (1) the Epidaurus Theatre in Greece without skenae wall or columnade behind the cavea; (2) the Jerash South theatre in Jordan with skenae wall but no columnade; and (3) the Aspendos Theatre in Turkey having both a full skenae building and a columnade around the cavea. In the paper the acoustic characteristics of these theatres will be compared and discussed in the light of acoustic performance requirements for modern use.

8:15

3aAAb2. Acoustical design issues for contemporary places of worship. Jason R. Duty and David R. Schwind (Charles M. Salter Assoc., Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94104, jason.duty@cmsalter.com)

This presentation intends to discuss the acoustical design of contemporary places of worship in the range of 500 to 3000 seats. The wide variety of programming requirements, concurrent uses, space constraints, and undesirable adjacencies in contemporary churches can be a similar challenge to many performing arts spaces. Many contemporary churches have loud, amplified music in their services and children's classrooms. In addition, religious study courses in adjacent classrooms are being held at the same time as the main worship service. During off-hours the spaces are used concurrently for teaching, conferences, weddings, prayer, and rehearsals. In many cases, the rooms outside of the main worship space are undersized and have undesirable adjacencies to each other as well as the main worship space. Using case studies we will discuss sound isolation design to reduce interference between concurrent unrelated uses, as well as room shaping and room finish treatments (including variable acoustics) to address room acoustics and the large sound systems.

8:30

3aAAb3. Acoustical design of Sejong Performing Arts Center recital hall. Jin Yong Jeon and Yong Hee Kim (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr)

Acoustical design of a convention center into a recital hall of Sejong Performing Arts Center located in Seoul, Korea was carried out to renovate its function. The area of the convention space is about 612 m^2 and it is mainly used for meetings, symposia, and so on. The volume of the existing hall is limited to 3100 m^3, which causes lack of late reflections, due to its structural limitations. Therefore, the existing ceiling slab was designed to be pulled up so that a new ceiling can create extra volume of 1000 m^3. A leaf-shaped floor plan and tilted side balconies within 16-m width between side walls were designed to adequately increase enough lateral reflection and spatial impression. The inclined side wall of 1:10 slope was proposed through the comparison with different angles of walls. The saw-tooth-shaped GFRC (glass fiber reinforced concrete) wall reflectors and center protruded stage rear wall are designed for sound diffusion. From the results of computer and scale modeling, it was predicted that the new space will have adequate reverberation time of 1.5 s and a uniform distribution of sound-pressure levels within 3 dB.

8:45

3aAAb4. A study on the influence of geometrical characters to the acoustical performance in vineyard terraced concert halls. Yi-run Chen and Wei-hwa Chiang (Dept. of Architecture, Natl. Taiwan Univ. of Sci. and Technol., #43, Sec.4, Keelung Rd., Taipei, 106, Taiwan, d9413002@mail.ntust.edu.tw)

There are two distinctive features defining vineyard terraced concert hall: audiences surrounding the stage and vineyardlike seating arrangement. Besides that, people may find more differences than similarities within those limited built cases. This paper investigated the influence of geometrical characters to acoustical performance in vineyard halls using computer simulation. Through case studies of nine vineyard terraced halls, some geometrical characters were selected as research variables, namely volume, overall proportion, dimensions of key elements, seating capacity, arrangement of vineyard terraces, terraced side box shaping, and ceiling shaping. Acoustical parameters EDT, LF, C80, and G were analyzed by varying architectural features of a prototype developed based on the main geometrical characters of the nine halls. It has been found that raising terraced side boxes would help increase G and LF. For a constant room volume, increasing the number of terraces with sufficient height would raise the LF value in the nearby area mostly. Changing ceiling shape only had influence on the rear part of the front terrace.

3aAAb5. Alternatives to the traditional concert hall experience.
Bobby GibbsII, Jonas Braasch, and Ted Krueger (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

As relates to live performance, the purview of sound quality in architectural acoustics is mostly limited to concert halls. While the concert hall is perhaps the most common setting in which to hear live Western classical music in our present day, this was not always the case historically. As will be shown from analysis of published text as well as interviews with music critics, artists, and concert-goers, a variety of alternatives to the concert hall exists currently in the form of domestic settings, casual settings like coffee shops, as well as large outdoor settings like farms and parks. The advantages and disadvantages of such alternative venues, as concerns the overall concert experience and the quality metric of intimacy particularly, will be discussed. Moreover, design issues relating to the acoustical environment as well as the social environment in these venues will be discussed based on the data. This paper will ideally foster discussion in the acoustical and the architectural community on ways to expand the traditional concert hall experience as well as address design issues in alternative venues. [Work supported by ASA Minority Fellowship and RPI Humanities, Arts, Architecture, and Social Sciences Fellowship.]

3aAAb6. New approach of the model experiment in room acoustics—advantage of $\frac{1}{16}$ scale model. Hidemaro Shimoda (Shimizu Corp., 3-4-17 Etchujima, Koto-Ku, Tokyo, 135-8530 Japan) and Yasuhiko Tahara (Tohoku Bunka Gakuen Univ., Aoba-Ku, Sendai, 981-8551 Japan)

In the construction of the auditoria, the examination by the acoustic model experiment has been often performed. This paper describes the outline of the result of investigating various features of the method of the experiment by $\frac{1}{16}$ scale models compared with models of $\frac{1}{10}$ and $\frac{1}{20}$ scales that have been generally adopted up to now. That is, the following advantages are found in the case study of the $\frac{1}{16}$ scale model experiment, and the new approach to the model experiment method is proposed here. (1) full cover of the band to 4 kHz enables the quantification of a practical physical acoustic parameters. (2) The parameter as the air absorption coefficient in the supersonic wave band by $\frac{1}{10}$ scale models can be used. (3) Cheap digital sound recording system can be used considering the microphone characteristics. (4) The model construction cost can be largely saved compared with $\frac{1}{10}$ models.

THURSDAY MORNING, 30 NOVEMBER 2006

LANAI ROOM, 9:45 TO 11:30 A.M.

Session 3aAAc

Architectural Acoustics: Acoustic Design Factors for Performing and Recording Spaces

David L. Adams, Cochair
David L. Adams Associates, Inc., 1701 Boulder St., Denver, CO 80211

Kanako Ueno, Cochair
Univ. of Tokyo, Inst. of Industrial Science and Technology, 4-6-1 Komaba, Meguroku, Tokyo 153-8505, Japan

Chair's Introduction—9:45

Contributed Papers

3aAAc1. Quantitative evaluation of characteristics of an auditorium shape in horizontal cross section. Kazuma Hoshi (Grad. School of Sci. and Technol., Nihon Univ., 1-8-14, Kandasurugadai, Chiyoda-ku, 101-8308, Tokyo, Japan, hoshi@arch.jcn.nihon-u.ac.jp), Toshiki Hanyu, and Katsuaki Sekiguchi (Nihon Univ., Chiyoda-ku, 101-8308, Tokyo, Japan)

The room shape of an auditorium is often classified as a "shoe-box type," "fan-shape type," "round type," "surrounding (vineyard) type," or other shape. The surface diffusivity index (SDI), which is an evaluation index of surface irregularities, has been proposed by Haan and Fricke, but the SDI is based on visual surveys of photographs or drawings [Appl. Acoust. **51**(1), 53–69 (1997)]. Clarification of the relationship between an auditorium's shape and its acoustic characteristics would be simplified if the auditorium shape were understood objectively rather than subjectively. Therefore, in this study, we tried to objectively obtain characteristics of a room shape in horizontal cross section using an elliptic Fourier descriptor, which can detect the periodicity of a two-dimensional contour. Room shapes were expressed as spatial frequency characteristics calculated from their Fourier descriptors. Based on them, we can understand the amplitude of irregularities and their scale. In addition, by using principal component analysis, it is possible to understand, quantitatively, the main differences of room shapes in horizontal cross section.

3aAAc2. Design and construction of a small sound testing room in an office building. Edward C. Duncan and Kenneth Kaliski (Resource Systems Group, Inc. 55 Railroad Row, White River Junction, VT 05001)

A small sound-testing room has been constructed in the basement of an office building located within 50 ft. of railroad tracks. The room was designed to be multifunctional with both reverberant and semianechoic settings. The sound isolation of the room was designed with the expectation of conducting active noise reduction tests with sound-pressure levels of 120 dB and higher. The room is also being used to roughly estimate absorption coefficients of various materials. The room is $10.5 \times 9.75 \times 7.5$ ft. with a volume of 768 ft^3. It is composed of a concrete floor and gypsum walls and ceiling. The walls are isolated from the floor with rubber isolation pads. The ceiling is constructed of 5/8-in. gypsum, 1/2-in. plywood, and a 2-in. layer of sand. The wall construction is 2×4 wood studs, two layers of 5/8-in. gypsum resiliently attached, and 3 in. of fiberglass. The entrance to the room is made of double solid core wood doors with a composite isolation and absorption layer in between both doors. Other design components, actual sound transmission testing results, reverberation times, modal issues, and uses for the room are discussed.

10:20

3aAc3. Introduction of auditory filter into evaluation of characteristics of small enclosures. Saori Shimada, Hisaharu Suzuki, and Akira Omoto (Omoto Lab., Dept. of Acoust. Design, Faculty of Design, Kyushu Univ., Shiobaru 4-9-1, Minami, Fukuoka 815-8540, Japan)

Physical measures, which are useful for evaluating small enclosures such as studios and car cabins, are examined. Especially, auditory filtering using Zwicker's loudness model is introduced for calculating the measures. Most current measures used in evaluating the large sound field such as concert halls are based on measured or calculated impulse responses. Instead, the music signals of some genres are used in this study. Using the auditory model, the amount of information can be reduced, even when complicated signals are used as stimuli: the signals are convertible to information in 25-bark bands. Correspondence of such physical quantities and subjective judgments are considered from the viewpoint of sound-field similarity. Subjects are asked to judge the similarity of two successive sounds whose frequency characteristics are modified artificially using the equalizer. The analysis results obtained using the auditory filter were compared with the subjective evaluation. Results indicate that analysis using the auditory filter can yield fairly good correspondence with the subjective evaluation, especially in a middle frequency band. Regarding the low-frequency band, the degree of correspondence depends on the musical genre. Expanding the effective frequency bands is an important theme of our ongoing research.

10:35

3aAc4. Cross-modal perception of room acoustics. Anthony McCreery and Paul Calamia (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, mccrea@rpi.edu)

The development of perceptually accurate audio-visual systems has been hindered by the fact that traditional scientific investigations into auditory and visual perception in humans have separated the two senses into separate and distinct modalities. It is important to consider the perception of these modalities as critically linked and ultimately inseparable. Research investigating expected levels of reverberance given 2-D visual representations of several environments shows that perception of reverberation is heavily dependent on the visual appearance of a room. For the experiments to be discussed here, subjects were asked to adjust the level of reverberant sound that accompanied video of a person speaking in three different rooms: a library, a playhouse, and a cathedral. The results show that reverberance levels chosen by subjects rarely match those measured from impulse responses recorded in the rooms being presented. Experimental methodology and results as well as implications for the development of future audiovisual systems will be discussed.

10:50

3aAc5. Considering source directivity in a recital hall with seats surrounding the platform. Wei Lin, Weihwa Chiang, and Yi-run Chen (Dept. of Architecture, Natl. Taiwan Univ. of Sci. and Technol., #43, Sec.4, Keelung Rd., Taipei, 106, Taiwan, d9313001@mail.ntust.edu.tw)

The impact of source directivity can be significant for chamber music performances, in particular for arena-type halls that have visual intimacy. This study has been carried out to understand the physical and perceptual aspects of the seats surrounding the platform in recital halls. Effects of architectural features were evaluated using computer simulations that accounted for source directivity and diffusive reflections. Both live and recorded subjective evaluations of chamber music performances were made in an actual hall while acoustical measurements were also taken. As a result, computer simulations using a singer source showed significant decrease in both clarity and strength in the 2-kHz band for the seats surrounding a centered platform. A wall 6 m in front of the platform was effective in mitigating the problems. Sound levels calculated from the recordings taken from various viewing angles were correlated with subjective evaluations, especially in the high frequencies. Timbre appeared to be the principal factor influencing solo performances and balance was the most important in judging ensembles. Nevertheless, live evaluation showed no significant difference among the listening positions.

Contributed Poster Paper

Poster paper 3aAc6 will be on display from 8:00 a.m. to 11:30 a.m. The author will be at the poster from 11:05 a.m. to 11:30 a.m.

3aAc6. Acoustic design of studio to record effect sounds in. Shinichi Oda, Junko Yokoyama, and Junichi Mikami (Dept. of Eng. Administration, Japan Broadcasting Corp. (NHK), 2-2-1 Jinnan, Shibuya-ku, Tokyo 150-8001, Japan)

We reconstructed a studio that has a space in which to record effect sounds. In this space, we record walking sounds on wooden floors, tile floors, sand, grass, and so on. This studio was planned on a floor with a light allowable load because the floor was formerly the roof of the building. A recording studio is usually constructed with a floating structure. The thin floating floor causes vibration of floorboards. A thick floor is heavy. In addition, we recorded the sound when we jumped on the floor. The noise from the structure disturbs the effect sounds that we need if the vibration caused by jumping stimulates the studio structure. For those reasons, we planned the following design to record pure effect sounds: The effect sound space is insulated from other spaces with the room structure. A floating ceiling is hung from the fixed ceiling, and parts of floating walls are hung from overhead beams. Only the floating floor and the other parts of floating walls are set upon the fixed floor. Consequently, the floating floor in the effect sound space was made thicker than would be possible otherwise.

Session 3aAB

Animal Bioacoustics: Acoustic Scene Analysis and Communication in Animals

Cynthia F. Moss, Cochair

Univ. of Maryland, Dept. of Psychology, Inst. for Systems Research, College Park, MD 20742

Hiroshi Riquimaroux, Cochair

Doshisha Univ., 1-3 Tatara Miyakotani, Kyotanabe, Kyoto 610-0321, Japan

Invited Papers

7:45

3aAB1. Auditory streaming and the spectrotemporal analysis of sound in the cortex. Shihab Shamma and Mounya Elhilali (Elec. and Comput. Eng. & Inst. for Systems Res., Univ. of Maryland, College Park, MD 20742)

Auditory streaming is a phenomenon that manifests itself in the everyday ability of humans and animals to parse complex acoustic information arising from multiple sound sources into meaningful auditory streams. For instance, the ability to follow a conversation at a noisy cocktail party or hear the violin in the orchestra both rely on the formation of auditory streams. While seemingly effortless, the neural mechanisms underlying auditory streaming remain a mystery. In this talk, we shall discuss how this perceptual ability may emerge as a consequence of the multiscale spectrotemporal analysis of sound in the auditory cortex. A simplified model of this process is developed to demonstrate how auditory streaming could underlie the separation of complex sounds from two sources (e.g., speech or music). The model reveals that certain perceptual attributes are critical for the simultaneous separation of sounds (pitch and onset sensiticity), whereas others are important for the sequential binding of the segregated sound features (timbre and location).

8:05

3aAB2. Time-sharing acoustic stream analysis to recognize multiple objects in space by echolocating bats. Hiroshi Riquimaroux (Doshisha Univ., 1-3 Tatara Miyakotani, Kyotanabe, Kyoto 610-0321, Japan)

Echolocating bats need to catch prey while they have to avoid obstacles and escape from their predators at the same time. Therefore, they have to perceive precise information about multiple objects from their surrounding space while they are flying. However, there are strong restrictions in the system; the neural system is too slow to perform real-time processing, and the bat has only one sound source emission and only two ears to receive echoes. The purpose of the present study was to investigate how echolocating bats perceive important information about different objects in space during their flight by analyzing acoustic characteristics of emitted pulses and returning echoes. Pulses and echoes of CF-FM and FM bats were recorded by a wireless telemetry microphone system (Telemike) on their head. In order to reconstruct their flight trajectory, a dual high speech video camera system was used. Obtained data showed multiple streams for temporal change in echo delay, Doppler-shift compensation, and echo intensity compensation. The present findings indicate that echolocating bats may adopt a time-sharing analysis, which is practically fast enough to recognize multiple objects in space during flight. [Work partly supported by the Innovative Cluster Creation Project promoted by MEXT and by a grant to RCAST at Doshisha University from MEXT.]

8:25

3aAB3. Echolocating bats control the directionality of sonar signals to sort out the auditory scene. Annemarie Surlykke (Inst. of Biol., Univ. of Southern Denmark, DK-5230 Odense M., Denmark) and Cynthia F. Moss (Univ. of Maryland, College Park, MD 20742)

A primary function of hearing is the localization and identification of sound sources in the auditory scene. Results from humans and other animals have shown that spectral and temporal features, as well as source direction, are important parameters used for organizing acoustic sources. Echolocating bats control the features of vocalizations to shape information carried by echoes from auditory objects. Our research on big brown bats, *Eptesicus fuscus*, suggests the importance of vocal control for scene analysis: When a bat performs a dual task, requiring both obstacle avoidance and prey capture, it solves the task sequentially, adapting the temporal pattern of its cries to produce sonar strobe groups, strobing first the obstacle and then the prey. Directional control over the sonar beam pattern confirms this sequential processing: The bat alternates fixation of the beam axis on the obstacle and prey. The importance of directionality is corroborated by studies in another bat, Trachops cirrhosus, where behavior and morphology optimize sonar directionality for different tasks. When hanging from a perch *T. cirrhosus* emits sounds for general orientation with limited demands for directionality. When hunting, prey localization necessitates high directionality associated with sound emission through the nostrils surrounded by the moveable nose leaf.

3151   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3151**

3a THU. AM

3aAB4. Wild dugongs' vocal responses to conspecific calls. Kotaro Ichikawa, Nobuaki Arai (Biosphere Informatics, Grad. School of Kyoto Univ., Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan, ichikawa@bre.soc.i.kyoto-u.ac.jp), Tomonari Akamatsu (Natl. Res. Inst. of Fisheries Eng., Kamisu, Ibaraki, 314-0408, Japan), Tomio Shinke (System Intech Co., Ltd., Shizuoka, Shizuoka, 424-8610, Japan), and Kanjana Adulyanukosol (Phuket Marine Biological Ctr., 83000, Phuket, Thailand)

Wild dugongs were found to call back more to conspecific calls than to artificially synthesized sounds. The population was exposed to four different playback stimulii a wild dugong chirp, a synthesized down-sweep sound similar to the dugong chirp, a synthesized constant-frequency sound, and no sound as a control. Vocalizing dugongs were localized using an array of stereo-underwater-recording systems. Wild dugongs vocalized more frequently after the playback of dugong chirps (2.8 calls/min) than those of constant-frequency (0.55 calls/min) and control (0.2 calls/min), ($p > 0.01$, Kruskal-Wallis test). Dominant frequencies of response calls were 4810 Hz to dugong chirps and 4470 Hz to down-sweep sounds. These were higher than those to other stimuli (3794 and 4044 Hz). Distances of calling-back dugongs from the playback speaker were significantly shorter for dugong chirps (10.19 m) and down-sweep (19.02 m) than that for constant frequency (105.84 m) ($p > 0.001$). The observed dominant frequencies of response calls (4510 Hz) were above the cutoff frequency of Lloyd's mirror effect in shallow waters of the present study. Frequency-modulated narrow-band sounds like chirps travel longer and enable accurate measurements of source directions by binaural receivers. These suggest that chirps could be used for conspecific recognition in noisy acoustic scenes in tropical shallow waters.

3aAB5. How does a green leafhopper male induce a female's willingness to mate? Masao Fukui (Lab. of Insect Physiol.., Grad. School of Agriculture, Kyoto Univ., Sakyo-ku, 606-8502, Japan)

When on their host plants, male leafhopper *Nephotettix nigropictus* (Insecta; Hemiptera) produces substrate-borne signals of two kinds, which are associated with different behavioral activities: (*i*) a calling signal and (*ii*) a duetting signal in response to a female. These signals both influence pair formation and the female's willingness to mate. After receiving a vibrational reply from a female responding to the calling signal, the calling male increases the first component of the calling signal from one trill to two or more trills, then continues in a duet with the female. During more duets, males wanting females to mate increase the number of chirps that compose a buzz as the second component of the calling signal. This increment of the first component is almost always one stable trill, but the increment of the second component apparently depends on the strength of the female's reply. This paper presents discussion of the biological significance and role of male signals to induce females to mate.

Contributed Papers

3aAB6. An avian cocktail party: Masking and unmasking in birds. Micheal Dent, Elizabeth McClaine (Dept. of Pyschology, B76 Park Hall, Unversity at Buffalo, Buffalo, NY 14260, mdent@buffalo.edu), Barbara Shinn-Cunningham, Virginia Best, Erol Ozmeral, Frederick Gallun, Kamal Sen, and Rajiv Narayan (Boston Univ., Boston, MA 02215)

The ability to effectively communicate in a noisy environment is an auditory skin that all animals must develop for basic survival. Birds are one group of animals that use auditory communication for everything from mating to prey capture. In general, how well humans and other animals are able to hear signals in their environment is a function of the properties of the signal and the properties of the noise, and include spatial, temporal, and spectral characteristics. Here, the abilities of birds to correctly identify zebra finch songs embedded in different types of maskers that were emitted from the same location or from separate locations (and presented at different signal to noise ratios) were measured using operant conditioning techniques. In general, zebra finches and budgerigars showed very small differences in thresholds between songs embedded in the three types of maskers: broadband noise, modulated noise, and choruses of zebra finch songs. When the calls were emitted from one location and the maskers from another, a condition that elicits spatial unmasking in humans (Best *et al.*, 2005), zebra finches had lower thresholds while budgerigars generally did not. These results suggest some species-specific advantages to listening in noisy environments.

3aAB7. Assessing the relative importance of source distance determined by acoustic cues and noise level as predictors of disturbance in a study of the Mexican spotted owl. Ann E. Bowles, Stefanie K. Moffat, Samuel L. Denes (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109, abowles@hswri.com), Christopher M. Hobbs (Wyle Labs., Arlington, VA 22202), and Dana H. Banwart (GeoMarine, Inc., Hampton, VA 23666)

In forested areas, acoustic localization of source distance and noise level are potentially reliable indicators of an approaching source. The relative importance of these cues was examined during a study of Mexican spotted owl responses to human-made noise. The study was conducted in mixed-conifer forest on steep northeast facing slopes. Owls had no or limited experience with the experimental paradigm. Shaped noise bursts were projected from a Sound Projections portable stereo speaker system with line of sight to the target owl(s). Signals were shaped noise bursts with 10–15-dB/s onset rate projected to produce received maximum A-weighted levels in two ranges, 55–65 dB and 75–85 dB, at distances of 20, 40, and 80 m. Levels were measured a short distance from the target owl(s) using a calibrated recording system (DAT or digital recorder equipped with an ACO 7013 microphone, 20 Hz to 20 kHz). They were also estimated using a single-event sound propagation model (NMSIM,Wyle Laboratories). Preliminary analysis showed that both source level and distance were significantly related to owl response, but that distance explained twice as much variance in behavior as level. [Work supported by ACC/CEVP; USFWS Permit No. TE024429.]

10:10

3aAB8. Comparisons of song perception in male and female songbirds utilizing cardiac responses. Maki Ikebuchi (Human Information System Labs., Kanazawa Inst. of Technol., Hakusan, Ishikawa 924-0893, Japan, mikebuch@his.kanazawa-it.ac.jp) and Kazuo Okanoya (RIKEN Brain Sci. Inst., Wako, Saitama 351-0198, Japan)

In most songbirds, males sing to defend territories and to attract females, and females evaluate the quality of males based on songs. This should lead to differences in song perception between the sexes, but it has been difficult to devise an assay that can evaluate song perception without introducing motivational bias. For example, the copulation solicitation assay is useful for evaluating female perception of songs, but this assay cannot be utilized for males. We found that birds react to a sound of interest by increasing the heart rate, and this response was used to compare song perception in Bengalese and zebra finches. We used conspecific and hetero-specific songs as stimuli and these were presented in a planned order. In Bengalese finches, only females responded to the changes in conspecific song repertoire. In zebra finches, both males and females reacted to such changes. Bengalese finches are a domesticated species and males do not establish breeding territories. Conversely, male zebra finches establish a small territory when breeding. In both species, songs are an important trait by which females select males. These ecological differences concur with the species and sex differences in the cardiac response. [Work supported by JSPS and JST.]

10:25

3aAB9. Detection of objects in complex environments by echolocating big brown bats. Caroline M. DeLong, Sarah A. Stamper, and James A. Simmons (Dept of Neurosci., Brown Univ., Box 1953, Providence, RI 02912)

In previous psychophysical experiments, big brown bats detected or discriminated objects (monopole and dipole targets comprised of 15 mm diameter cylinders) presented standing on smooth surfaces with little clutter. Performance was determined by mutual masking of echoes from the targets themselves. In new experiments, bats did two-alternative (left/right) forced choice tasks to detect a two-cylinder dipole target in complex, cluttered environments. In experiment 1, the targets were set inside 14-mm-deep, 20-mm-diam holes in a layer of foam 25 mm thick. Cylinders of different heights protruded 13, 5, or 2 mm above the surface of the foam, or were recessed 1 mm below the surface. The bats performance varied as a function of protruding height. In experiment 2, the dipole target was embedded within arrays of distractor objects that varied in shape, size, orientation, and material. Both experiments explore how bats isolate the echoes from the dipole target from the mixture of echoes returning from the clutter—whether distractor objects or echoes from the sides of the holes. Big brown bats must engage in this acoustic scene analysis when catching insects, such as beetles, against a backdrop of foliage.

10:40

3aAB10. The sonar beam of big brown bats (*Eptesicus fuscus*) during landing. Jens C. Koblitz, Peter Stilz, Wiebke Pflästerer, Mariana Melcon, and Hans-Ulrich Schnitzler (Tierphysiologie, Universität Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany, jens.koblitz@web.de)

The parameters signal duration, pulse interval, and frequency have previously been used to describe the approach behavior of landing bats. An array with 16 microphones was used to investigate additional parameters during the approach flight of landing big brown bats (*Eptesicus fuscus*). For each signal emitted the position of the landing bat in relation to the microphone array was determined using a 3-D video system, and the width, maximum intensity, and aiming direction of the sonar beam were calculated. The approach to the landing site begins at a distance of about 1.4 m when the bats pick up the landing site with their sonar beam. This focusing on the landing site is indicated by a clear decrease of the deviation between beam aiming and direction to the landing position. The maximum intensity of the sonar beam is reduced by about 6 dB per half distance during the last 2 m of the approach, which indicates a gain control mechanism. The width of the sonar beam remains constant throughout the whole approach.

10:55

3aAB11. Spatial release from masking of aerial tones in a California sea lion (*Zalophys californianus*). Marla M. Holt and Ronald J. Schusterman (UCSC Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060)

Spatial release from masking (SRM) occurs when a signal and masker are spatially separated, resulting in improvement of signal detection relative to when they are spatially co-located. Sea lions forage in the water, breed on land, produce airborne vocalizations that are associated with social and reproductive activities, and possess highly reduced pinnae. In this study, SRM was measured at 1, 8, and 16 kHz in a California sea lion who had to detect an aerial tone in the presence of an octave band of white noise centered at the tone frequency. Testing was conducted in a hemi-anechoic chamber. While the masker always occurred in front of the subject (0 deg), the tone occurred at 0, 45, or 90 deg in the horizontal plane. Absolute thresholds were also measured at these angles to account for differences in hearing sensitivity based on source azimuth. Masked thresholds were lower by as much as 12 dB at 1 kHz when the signal and masker were separated by 90 deg. These results were compared with those of a harbor seal who, like all true seals, naturally lacks pinnae. Performance differences between the two subjects at the highest frequency likely reflect variations in pinna anatomy.

11:10

3aAB12. Underwater noise and zones of masking with respect to hopper dredging and manatees in the St. Johns River in Jacksonville, FL. Edmund R. Gerstein, Joseph E. Blue (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), Gerard F. Pinto, and Seth Barr (Jacksonville Univerisity, Jacksonville, FL 32211)

Underwater noise radiating from dredging can effectively obscure or mask biological and other important sounds. This study recorded underwater acoustic characteristics of hopper dredging in the St. Johns River, Jacksonville, FL, to evaluate noise impacts in the waterway with respect to the endangered West Indian manatee. Of particular interest was the extent and range that dredging noise may mask the sounds of approaching commercial and recreational vessels. Vertical hydrophone arrays and a multichannel PC-based recording system were used to measure dredging noise at various distances. Ambient noise surveys, active propagation of calibrated sources, and controlled boat noise measurements were conducted along the waterway. These data were integrated with behavioral hearing data to estimate zones of masking surrounding dredging. Three discernable noise sources that masked boat noise were (1) cavitation from dredge propellers, (2) draghead vacuuming, and (3) noise from submerged slurry pipelines. Sustained high ambient noise levels from dredging can significantly increase the risk of manatee-boat collisions by masking the sounds of approaching vessels over large radii (up to 2.5 miles). Mitigations suggested include ship quieting, reducing propeller cavitation, insulating or elevating slurry pipelines, and minimizing transects to pump out stations. [Work funded by the City of Jacksonville Waterways Commission.]

3a THU. AM

3aAB13. "Vocal grooming" in an eusocial subterranean rodent: The naked mole-rat (*Heterocephalus glaber*). Shigeto Yosida, Kazuo Okanoya (Grad. School of Sci. & Technol., Chiba U., 1-33 Yayoi-Cho, Inage, Chiba, 263-8522 Japan), Chizuko Suzuki, Sanae Nanvuh, and Kazuo Okanoya (RIKEN BSI., Wako, Saitama, 351-0198 Japan)

The naked mole-rat is a unique communal-living rodent of east Africa. In a subterranean environment, this small rodent species has a complex hierarchical society with labor division that is comparable to those of bees and ants. They are also known for their extensive vocal repertory of more than 17 sounds, which might have evolved as an alternative to visual cues, and which contributes to maintaining its society. We showed that their most frequent vocalization, the soft-chirp (SC), has an antiphonal nature, and it might function for rank-identification and reconciliation purposes. The SC is elicited either by another's SC or physical touch. We analyzed acoustic properties of SC elicited by touching the animals, and found that it varies according to the body mass. The emission rate was highest among larger (superior) animals, but the smaller (inferior) of the two emitted SCs more frequently when observing behavior of two interacting animals. Playback experiments revealed that the use of SC generally follows these two rules: smaller animals vocalize more often than larger ones, and larger ones respond only to smaller animals' SCs. These features of SC resemble grooming behavior among primates.

3aAB14. Neural responses to communication sound in the inferior colliculus of the naked mole-rat. Shizuka Tokin, Kohichi Tanaka (Lab. for Molecular Neurosci., Sch. of Biomed. Sci., Tokyo Med. and Dent. Univ., 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-0034 Japan), Shizuka Tokin, and Kazuo Okanoya (RIKEN BSI., Wako, Saitama, 351-0198 Japan)

The naked mole-rat (*Heterocephalus glaber*) is a highly social subterranean rodent that lives in Eastern Africa. Because they a dig burrow system and spend most of their lives underground, they are functionally blind and rely heavily on auditory information. They have at least 17 different vocalizations, which are associated with specific behavioral contexts. In this study, we focused on the soft chirp (SC), which is one of the most frequent vocalizations in this species, and recorded neural responses to synthesized SCs in the inferior colliculus. We found that the responses could be divided into three groups based on their response patterns. These response patterns represented the duration or frequency of stimuli. In addition, we also found that the number of spikes changed with the degree of frequency modulation. From these results, we hypothesized that the vocalizations of naked mole-rats can be represented by integrating the activity of such neurons that respond to a specific feature of sound.

THURSDAY MORNING, 30 NOVEMBER 2006　　　　　MAUI ROOM, 8:00 A.M. TO 12:00 NOON

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Celebration of Floyd Dunn I

William D. O'Brien, Cochair

Univ. of Illinois, Dept. Electrical and Computer Engineering, 405 N. Mathews, Urbana, IL 61801

Junichi Kushibiki, Cochair

Tohoku Univ., Dept. of Electrical Engineering, 6-6-05 Aoba, Aramaki-Aza, Aoba-ku, Sendai 980-8579, Japan

Chair's Introduction—8:00

Invited Papers

8:05

3aBB1. Celebration of Floyd Dunn: Some remarks about his life, career, and accomplishments. William D. O'Brien, Jr. (Bioacoustics Res. Lab., Dept of Elec. and Comput. Eng., Univ. of Illinois, 405 N. Mathews Ave., Urbana, IL 61801)

No celebration of Floyd Dunn would be complete without some remarks about the life, career, and accomplishments of this remarkable scientist/engineer. Floyd had made significant scholarly contributions in six general themes, and you will hear about their impact to our welfare during the celebration. The six ultrasonic biophysics themes include absorptive processes, nonlinear phenomena, application in living systems, toxicity, measurement techniques, and ultrasonic microscopy. Floyd has been recognized for his accomplishments by being the recipient of most of the important national and international awards including National Academy of Sciences, National Academy of Engineering, ASA's Gold Medal and Silver Medal in Bioresponse to Vibration, Acoustical Society of Japan's Medal of Special Merit, IEEE's Edison Medal, and election to Fellowship in six professional societies.

8:25

3aBB2. Imaging propagation speed of pulsive vibration in the heart wall for healthy subjects and patients with old myocardial infarction. Hiroshi Kanai (Dept. of Electron. Eng., Tohoku Univ., Aramaki-aza-Aoba 6-6-05, Sendai 980-8579, Japan), Sachiko Watanabe (Miyagi Shakaihoken Hospital, Sendai, Japan), Yoshifumi Saijo, and Motonao Tanaka (Tohoku Univ., Sendai, Japan)

This study presents a novel method that noninvasively visualizes myocardial heterogeneity for healthy subjects and patients with old myocardial infarction. By a sparse sector scan, the myocardial minute motions were measured simultaneously at about 1000 points in the heart wall at a high frame rate. We have found that a pulsive vibration is excited by aortic-valve closure [H. Kanai and Y. Koiwa, Ultrasound Med. Biol. **27**, 481–498 (2001)]. The consecutive spatial distributions of the vibrations reveal that the vibration propagates from the root of the aortic valve to the apex [H. Kanai, IEEE Trans. UFFC. **51**, 1931–1942 (2005)]. In this study, the

method was applied to seven healthy subjects and five patients, and the spatial distribution of the propagation speed of the vibration is also visualized. For all healthy subjects, the propagation of the pulsive vibration was visible both in longitudinal-axis and apical views, and there is spatial difference in the propagation speed; 4.3 ± 0.5 m/s for the high echo area and 2.9 ± 0.4 m/s for the low echo area, for a 60-Hz component. For patients, the propagation speed in the diseased area is about 30% lower than the normal area. These preliminary results indicate that the method has a potential for revealing myocardial heterogeneity.

8:45

3aBB3. Ultrasound stimulated vibroacoustography and vibrometry for measuring tissue properties.   James F. Greenleaf and Chen Shigao   (Mayo Clinic College of Medicine, Rochester, MN 55901)

Ultrasound stimulated vibroacoustography has been used to image tissues using the acoustic response of tissues to localized harmonic radiation pressure. The method provides high resolution and high dynamic range images of tissues. The parameter being imaged is a complex combination of scattering, attenuation, and nonlinearity. However, specific use of harmonic or pulsed radiation pressure and subsequent measurement of the tissue response can be used to measure fundamental material properties of tissue. The measurement of shear wave dispersion can be used to estimate elastic shear moduli of tissue. Fundamental model free properties such as elastic storage and loss constants can also be measured. Ultrasound radiation pressure is used to induce free propagating shear waves. The measurable properties of the shear waves are sensitive to only the material properties of the tissue under certain circumstances. Shear moduli in tissue are measured with high accuracy and precision given appropriate models of wave propagation within the geometry of the tissue. Careful use of shear wave propagation and subsequent measurements can provide fundamental measurements of tissue mechanical properties if models are accurate.

9:05

3aBB4. Evaluation of liver fibrosis using the acoustic characteristic and statistical signal processing of echo.   Tadashi Yamaguchi   (Faculty of Eng., Chiba Univ., 1-33 Yayoi-cho, Inage-ku, Chiba, 263-8522 Japan) and Hiroyuki Hachiya   (Chiba Univ., Inage-ku, Chiba, 263-8522 Japan)

We evaluate the relationship between changes in acoustic properties accompanying change of a tissue and echo information using human liver specimens *in vitro*. The rf echo signals of a 20-mm-thick human liver specimen (200 frames) were acquired continuously using a probe using 2.0-MHz transmission frequency and 4.0-MHz receiving frequency. The signals were processed using statistical signal processing to extract abnormal echo information. After ultrasonic scanning, 50×50-mm^2 sections were cut from the liver specimens for microscopy and acoustic microscope measurement. The liver fiber structure was extracted from 20 4-μm-thick Masson's trichrome stained microsections. The speed of sound and the attenuation of 2-mm-thick cut specimens were measured using 25-MHz ultrasound. Results of comparing 3-D structures of abnormal information extracted from echo signals and the 3D fiber structures extracted from microsections showed that both structures were clearly in agreement. Furthermore, it was confirmed that the speed of sound of fibers measured using an acoustic microscope was faster than that for normal liver tissue.

9:25

3aBB5. The importance of cells and arrangement of cells to ultrasound backscatter.   Michael Oelze   (Dept. of Elect. & Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801, oelze@uiuc.edu)

The early work of Professor Floyd Dunn identified the importance of collagen structures and globular protein to ultrasound backscatter [Fields and Dunn, J. Acoust. Soc. Am. **54**, 809–812 (1973)]. Recent advances in tissue characterization through ultrasound backscatter have again focused on determination of the underlying microstructures responsible for scattering. The focus of new models for ultrasound backscattering has been on the contribution of cells. Specifically, new models have sought to address the ultrasound backscatter from single cells and from different arrangement of cells. The importance of cells and their arrangements in tissues to classification of disease, i.e., cancer, through ultrasound was also noted by Professor Dunn [Saijo *et al.*, Ultrasound Med. Biol. **17**, 709–714 (1991)]. In the present work, the construction of a new cell model of scattering was made from measurements of acoustic properties of cells. Different arrangements of cells, i.e., uniform distribution of cells and cells arranged in groups, were examined in simulation using the new cell model. Finally, the new cell model and simulations were compared to ultrasound backscatter measurements from different kinds of cancer. The arrangement of cells within a cancerous region was observed to be significant for diagnostically differentiating between cancers using ultrasound backscatter.

9:45

3aBB6. Evaluation of blood flow velocity vector measurement accuracy: Numerical simulation and flow phantom studies of Rankine's vortex.   Takashi Okada, Teruyuki Sonoyama, Akimitsu Harada   (Aloka Co., Ltd., 3-7-19, Imai, Ohme, Tokyo, 198-8577, Japan), Shigeo Ohtsuki   (Tokyo Inst. of Technol., Nagat-suda, Midori-ku, Yokohama, 226-8503, Japan), and Motonao Tanaka   (Tohoku Employee's Pension Welfare Hospital, Hukumuro, Miyagino-ku, Sendai, 983-8512, Japan)

A new method to measure the velocity vector of blood flow on the observation plane was recently developed [Ohtsuki and Tanaka, J. Visuali. **9**, 69–82 (2006)]. The method is based on the calculation of stream functions. To evaluate the accuracy of this method, we have studied the velocity distribution of Rankine's vortex in numerical simulation and experiments. The vortex of the simulation model was 50 mm in radius, the forced vortex was 20 mm in radius, and the maximum velocity of the vortex was 50 cm/s. The differences between the velocity profiles of the vortex obtained by this method and by simulation were less than 10%, although the accuracy depended on the angle of the observation plane. Furthermore, we conducted phantom experiments in which the vortex was created in degassed water by using a magnetic stirrer. Color Doppler data of the vortex flow were acquired using a ultrasonic diagnostic system (SSD-6500, Aloka, Japan) and the velocity vector distribution was calculated. The experimental results were in good agreement with the simulation model. We conclude that the reliability of this method is very high. The results suggest great potential of the new velocity vector method for quantitative diagnosis of cardiac functions.

10:20

3aBB7. Professor Floyd Dunn's contribution to understanding the biological effects of ultrasound. Mel E. Stratmeyer   (Food and Drug Administration/CDRH, 9200 Corporate Blvd. (HFZ-120), Rockville, MD 20850, melvin.stratmeyer@fda.hhs.gov)

During Professor Dunn's career, ultrasound exposure has evolved from being used to induce dramatic tissue effects in animals to routine diagnostic fetal imaging in humans. An understanding of the fundamental mechanisms and dosimetric parameters involved in ultrasound-induced biological effects was essential to allow ultrasound to be clinically used for tissue destruction, physical therapy, and routine diagnostic imaging. Professor Dunn's early research and the research of other prominent scientists, many of them trained by Professor Dunn, will be discussed to demonstrate advancements in our understanding of the biological effects of ultrasound.

10:40

3aBB8. Reduction of cavitation threshold by second-harmonic superimposition and its potential application to sonodynamic therapy. Shin-ichiro Umemura   (School of Health Sci., Kyoto Univ., Kyoto, 607-8507 Japan, sumemura@hs.med.kyoto-u.ac.jp), Ken-ichi Kawabata, Kazuaki Sasaki, and Nami Sugita   (Hitachi Central Res. Lab., Tokyo, 185-0043 Japan)

Acoustic cavitation is known to induce bioeffects under certain conditions. Cavitation threshold, the acoustic strength needed to induce such cavitation, in various biological systems was intensively studied by Dr. Dunn's group. Their results have significantly contributed to the safety concerns of diagnostic and therapeutic uses of ultrasound. Acoustic cavitation is also known to be the primary cause of sonochemical reactions. Chemical compounds such as hematoporphyrin were found to be activated by acoustic cavitation and induce *in vitro* and *in vivo* antitumor effects thereby. In order to make use of this effect in therapeutic practice, cavitation must be generated reproducibly without depending on standing waves at acoustic power inducing no significant heating. It was also found that the cavitation threshold can be significantly reduced by superimposing the second harmonic onto the fundamental in insonation. The threshold was reduced by more than an order of magnitude, especially in the presence of a certain compound such as Rose Bengal (RB). An amphiphilic derivatives of RB, selectively accumulating in tumor tissues, was synthesized and used in combination with focused ultrasound with second-harmonic superimposition to treat experimental murine tumors, resulting in significant reduction of their sizes. A hypothesis on the mechanism behind this will also be discussed.

11:00

3aBB9. Exposure estimation and dosimetry for ultrasound thermal therapies. Gail ter Haar   (Joint Dept. of Phys., Inst. of Cancer Res., Royal Marsden Hospital, Sutton, Surrey, SM2 5PT, UK)

The success of ultrasound thermal therapies relies on achieving a predetermined temperature rise in a specified tissue volume. If these treatments are to be used in clinical application, it is important to be able to estimate the required exposure parameters and to undertake appropriate dosimetry. Floyd Dunn and his co-workers were pioneers in this field for high-intensity focused ultrasound (HIFU). In early papers they addressed the problem of ultrasonic dosage and dose rate. This remains a topic of active discussion. A major part of these investigations relies on accurate temperature measurement in ultrasonic fields in tissue. The viscous heating artifact that occurs when thermocouples are used was first recognized by the Illinois group in the 1950s. Effective ways of estimating this component of the measured temperature rise are still being sought. Investigation of the dependence of HIFU lesion size on *in situ* ultrasound intensity, exposure time, and frequency will be described, and the importance of acoustic cavitation and boiling for the definition of an ultrasound dose parameter will be discussed.

11:20

3aBB10. Numerical analysis of temperature in tissue radiated by focused ultrasound. Nobuyuki Endoh and Takenobu Tsuchiya (Dept. of Eng., Kanagawa Univ., Yokohama 221-8686, Japan)

Higher power ultrasonic equipment is recently extending not only into therapy but also into diagnosis because newer diagnostic systems such as harmonic imaging require higher ultrasound than conventional imaging systems. It is very important to assure the safety in the use of diagnostic equipment, especially for prevention of burns caused by radiation from ultrasonic imaging systems. In this two-dimensional FDTD-HCE study, temperature increases in tissue were simulated at the focal point of a phased array transducer like a common B-mode imagine system. A center frequency of radiated pulse is assumed to be 2.5 MHz and ISTPA = 0.72 W/cm^2. Results show that the temperature in tissue at the focal point is proportional to the pulse repetition frequency (PRF) when the PRF is changed from 100 to 400 kHz. It is also shown that common equipment is safe because the maximum temperature rise in tissue is only 0.0004 °C at the focal point of the transducer even though the PRF is 400 kHz. [Work supported by High-Tech Research Center Project from the Ministry of ECSST.]

11:40

3aBB11. Professor Floyd Dunn's contributions to ultrasound metrology. Gerald R. Harris   (U.S. Food and Drug Administration, 12725 Twinbrook Pkwy., Mail Code HFZ-170, Rockville, MD 20852-1719, gerald.harris@fda.hhs.gov)

A half-century ago the applications of ultrasound in medicine were, apart from physiotherapy, mostly experimental, and the related bioeffects and toxicity studies were in their infancy. Progress in both of these areas was hampered in part by the lack of acceptable tools for measuring the exposure levels, especially for the intense fields then being used in ultrasound surgical studies. Professor Floyd Dunn's advancements in ultrasound measurement techniques at the University of Illinois Bioacoustics Research Laboratory afforded accurate means for quantifying these exposures and laid the groundwork for subsequent developments in ultrasound measurement and calibration. In this presentation Professor Dunn's early work in thermal and radiation force techniques will be described and some recent methods that are being investigated to characterize ultrasound fields, particularly at high intensities, will be reviewed.

Session 3aEA

Engineering Acoustics, ASA Committee on Standards, and Psychological and Physiological Acoustics: Hearing Aids

Stephen C. Thompson, Cochair

Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Yasuhiko Arai, Cochair

Kanazawa Inst. of Technology, 7-1 Ohgigaoka, Nonoichi, Ishikawa 921-8501, Japan

Invited Papers

8:00

3aEA1. Digital hearing aids: A brief history.　Harry Levitt　(P.O. Box 610, Bodega Bay, CA 94923)

The journey leading to the development of digital hearing aids began at Bell Laboratories in the mid 1960s when a digital computer was used to simulate an experimental high gain telephone with frequency shaping for people with hearing loss. The simulation was off-line at about 100 times real time. About 20 years later, a digital master hearing aid operating in real time was developed using a high speed array processor. The development of wearable digital hearing aids followed soon after as high speed digital signal processing (DSP) chips were developed. The first wearable digital hearing aid was a body worn instrument and was not a commercial success. At this stage, digitally controlled analog hearing aids could be made small enough to be worn on the ear and these instruments were widely used by the end of the decade. During the 1990s, low-power DSP chips were developed that were small enough to allow for all-digital ear-worn hearing aids to be developed. The challenge today is to tap the tremendous potential of digital signal processing techniques within the power and size constraints of a practical, wearable instrument.

8:20

3aEA2. A new speech enhancement method for two-input two-output hearing aids.　Junfeng Li, Shuichi Sakamoto, Yo-iti Suzuki　(R. I. E. C., Tohoku Univ., 2-1-1, Katahira, Sendai, 980-8577, Japan, junfeng@ais.riec.tohoku.ac.jp), and Satoshi Hongo (Miyagi Natl. College of Tech., Medeshima-Shiote, Natori-shi, Miyagi, 981-1239, Japan)

Human beings have the ability to pick up a speech signal in noisy environments, which is known as the cocktail-party effect. This phenomenon is, however, often degraded in persons with impaired hearing. Therefore, a good method to enhance this phenomenon must be very useful to improve speech intelligibility in ambient noise for hearing-impaired persons. In this lecture, multi-input two-output speech enhancement techniques are first summarized. Subsequently, a new two-input two-output speech enhancement method is proposed that is based on the frequency domain binaural model (FDBM) [Nakashima *et al.*, Acoust. Sci. Technol. **24**, 172–178 (2003)]. In the proposed method, the interaural differences are first calculated from the noisy observations and employed to determine the sound-source locations. A speech absence probability (SAP) estimator is then developed using knowledge of estimated source locations. It is further integrated into the original FDBM to improve the interference-suppression capability. Effectiveness and superiority of the proposed speech enhancement method with respect to the original FDBM are validated through comprehensive experiments in terms of objective and subjective speech-quality measures. [Work supported by Sendai Intelligent Knowledge Cluster.]

8:40

3aEA3. Utilizing the auditory systems processing capabilities in the development of methods to auto-steer signal processing strategies in hearing instruments.　Andrew B. Dittberner, Jeff Bondy, Maureen Coughlin, Bill Whitmer　(GN Auditory Res. Lab., 2601 Patriot Blvd., Glenview IL 60026), and Melissa Dominguez　(DBK Acoust., Grayslake IL, 60030)

Hearing instruments have evolved in both the hardware and signal processing domains, providing more computational capabilities and sophisticated sound processing strategies. Static or adaptive methods are employed by these sound processing strategies to achieve a desired output. Due to the adaptive nature of some of these processing strategies and the continual increase in the number of parameters that can interact negatively with each other, there is a need to provide some method of global control over the parameter states. In addition, due to individual needs of the end user, parameter states in a hearing instrument may not satisfy the broad range of acoustic environments that a listener encounters, so identification of particular acoustic environment characteristics is also important. The purpose of this work is to introduce and discuss these problems and some of the current research efforts addressing environmental classification and global control of a hearing instrument's parameter states. The scope of this discussion will be on strategies employed to meet the needs of the end user in reference to the acoustic environment they identify as having a negative impact on their auditory experience and the desired performance they expect from the hearing instrument in that environment.

3a THU. AM

3aEA4. A study of signal-to-noise ratio conditions for measuring the intelligibility of hearing aids in noisy live environments.
Yasuhiko Arai (Kanazawa Inst. of Technol., 7-1 Ohgigaoka, Nonoichi, Ishikawa 921-8501, Japan, araiyasu@neptune.kanazawa-it.ac.jp) and Yoshiyuki Yoshizumi (Matsushita Electric Ind. Co., Ltd., Tsuzuki-ku, Yokohama 224-8539, Japan)

The authors studied a method to measure the intelligibility of directional hearing aids using a dummy head in noisy environments. The method is based on ISO recommendations (ISO 8353-2, ISO 8353-3). However, it is sometimes difficult to meet the quasi-free-field conditions. Therefore, a method for approximate measurement of intelligibility, by which the equivalent results are obtained in noisy live environments, is recommended. To simulate noisy environments, a multitalker (MULT) noise or a simulated multitalker noise was simultaneously and noncorrelatively put out through four loudspeakers in the live testing room. The MULT noise was composed of eight voices (four males and four females). The experiment showed that mixed voices of eight to ten persons were adequate for interference of the test speech words. The 50% SNR specification, meaning signal-to-noise ratio that gives intelligibility of 50% correct, was introduced into the measurement system. Intelligibility was measured under various SNR conditions. It was expressed approximately by the sigmoid function of SNR. The 50% SNR was estimated from the experimental equation. The intelligibility was measured adequately in the range between 0% and 100% by using the 50% SNR specification. The system configuration, measuring conditions, and measured samples are discussed.

3aEA5. Perception-model-based sound quality prediction for hearing aids: Big effect, small effect, and the optimization of noise reduction algorithms.
Birger Kollmeier, Rainer Huber, Thomas Rohdenburg, and Volker Hohmann (Medizinische Physik, Universitaet Oldenburg and Kompetenzzentrum HoerTech, D-26111 Oldenburg, Germany)

Objective sound quality models can be used to optimize hearing aid algorithms and their respective parameter settings by numerically predicting the subjective sound quality assessed by humans without having to perform quality measurements with humans concurrently. The big effect prediction considers how the (individual) hearing loss affects the aided or unaided perceived sound quality. The small effect prediction aims at subtle differences at a given amplification level of the hearing aid in order to compare processing parameters. This talk will concentrate on benchmarking different types of objective prediction methods both for big effect and small effect conditions using a wide range of signals and processing conditions, respectively. In addition, the performance of PEMO_Q, an auditory-model-based quality prediction scheme developed at Oldenburg University will be discussed, which has been modified to include the effect of hearing impairment and the adaptation to different kinds of test data sets. We will also demonstrate the degree to which PEMO_Q is beneficial for optimizing noise reduction algorithms that are suitable for hearing aids. The objective measure helps to efficiently scan the multidimensional parameter space for an optimum parameter set of the noise reduction algorithm. [Work supported by BMBF and by HoerTech.]

3aEA6. The design and evaluation of a noise suppression system for custom-made digital hearing aids.
Yoshiyuki Yoshizumi and Shigekiyo Fujii (Matsushita Electric Industrial Co., Ltd., 600, Saedo-cho, Tsuzuki-ku, Yokohama, 224-8539 Japan, yoshizumi.yoshiyuki@jp.panasonic.com)

A noise suppression system, which has the inbuilt feature to adapt to the local environment, for custom-made digital hearing aids has been developed. The signal process is based on spectral subtraction technology with 32 bands in the frequency domain. The signal-to-noise ratio was improved by an average of 7.5 dB in rms level. Its effectiveness was evaluated using the seven-point rating scale method for nine normal-hearing subjects and six hearing impaired subjects. Test speech is presented under three different environments, which consist of a car, a duct, and a crowded place. As a result of the assessment, the same tendency is observed between normal and hearing-impaired subjects and there were two good points that were found. First, the noise suppression system makes them more comfortable to listen to speech with the statistical significance of $p = 0.05$. Second, the noisiness is reduced with the statistical significance of $p = 0.01$. It should be noted that there is a scattering of the plotted data; however, statistically, the data result is quite acceptable.

3aEA7. Development of novel biologically inspired directional microphones.
Ronald N. Miles, Weili Cui, Quang Su, Rui Wu, Lin Tan, Yang Liu, Stephen A. Jones, Venkatesh Mohnankrishnaswami, Thomas Strait, William Butler, David DiBernardo (Dept. of Mech. Eng., State Univ. of New York, Binghamton, NY 13902-6000, miles@binghamton.edu), F. Levent Degertekin, Baris Bicen, Wook Lee, and Kamran Jeelani (Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The development of novel directional microphones for hearing aids is described. The mechanisms underlying the design of these unusual microphones were inspired by our earlier study of the ears of the parasitoid fly *Ormia ochracea* [Miles, *et al.*, J. Acoust. Soc. Am. **98**, 3059–3070 (1995)]. The structure of *Ormia's* ears inspired new approaches to the design of directional microphones that have the potential to be more sensitive and have lower thermal noise than typical miniature microphones. The mechanisms for directional hearing in this animal are discussed along with the engineering design concepts that they have inspired. Microphones have been fabricated out of silicon that employ either capacitive sensing or optical sensing to convert the diaphragm motion into an electronic signal. Measured results indicate that the directivity of these microphones is very similar to that of an ideal first-order differential microphone. In addition, novel microphone diaphragms have been fabricated that posses a second-order directional response. These can be used to achieve a significant reduction of unwanted background acoustic noise in hearing aid applications. [Work supported by NIH Grant 5R01DC005762-03, Sensing and Processing for Directional Hearing Aids, to RNM.]

Contributed Papers

11:00

3aEA8. Binaural speech enhancement by complex wavelet transform based on interaural level and argument differences. Satoshi Hongo (Dept. of Design and Comput. Applications, Miyagi Natl. College of Technol., 48 Nodayama, Medeshima-Shiote, Natori, Japan, hongo@miyagi-ct.ac.jp) and Yôiti Suzuki (Res. Inst. of Elec. Commun., Tohoku Univ., Japan)

Binaural information might enhance speech signals in noisy environments. Most precedent studies in this area have implemented signal processing in time and frequency domains. For this study, an enhancement method using complex wavelet transform (CWT) was proposed. The CWT has a scale domain whose bandwidth is inversely proportional to the scale level and may therefore be well compared to auditory filters. The proposed processing procedure is the following: (1) By computing CWT for a sound signal at every direction of arrival (DOA), a database (DB) of the scale domain wavelet coefficients is prepared for every DOA. (2) For binaural input signals from an unknown direction, the scale domain wavelet coefficients are calculated using CWT. (3) The DOA of the input signal is estimated as the direction for which interaural level and argument differences calculated from wavelet coefficients are the most similar to those in the DB. (4) The input signal is segregated similarly to FDBM (Nakashima *et al.*, 2003) based on the estimated DOA. Experiments with a directional target signal and an interference sound source calculated from real HRTFs were conducted to show the effectiveness of the proposed algorithm. [This research was partially supported by Sendai Intelligent Knowledge Cluster.]

11:15

3aEA9. Study of the multimicrophone composed of unidirectional microphones for glasses-type hearing aids. Kenzo Miura (Panasonic MCE Co., Ltd., 4-3-1, Tsunashima-higashi, Kohoku-ku, Yokohama 223-8639, Japan) and Yasuhiko Arai (Kanazawa Inst. of Technol., Nonoichi, Ishikawa 921-8501, Japan)

That multiunit microphones (MUMs) using unidirectional units offer high sensitivity, low electronic noise, and low wind noise for the binaural hearing aids of glasses type is a fact that has already been reported. [K. Miura and Y. Arai, Th.2.D.4, 18th ICA, Kyoto, Japan (April, 2004)]. Two types of MUM are proposed herein for glasses-type hearing aids that have an array of three units positioned over the front frame for each channel. One type is the parallel front-direction array type (PA-MUM). The other type has a twisted direction array (TA-MUM). Performances of these MUMs were compared with that of a bigradient (BG) microphone positioned on the side frame. Results show that the PA-MUM has very narrow directivity; TA-MUM showed moderate directivity. The electronic noise and the wind noise of these MUMs were at a very low level. On the other hand, the BG microphone showed rather sharp directivity, but the electronic noise from the signal conditioner and the wind noise were both of a high level. These results show that PA-MUM is suitable for listening to speech in very noisy environments and that TA-MUM is suitable for general use for glasses-type hearing aids.

11:30

3aEA10. A model of own-voice perception with hearing aid usage. Stefan Stenfelt (Dept of Neurosci. and Locomotion/Div. of Tech. Audiol., Linkping Univ., SE-58185 Linkping, Sweden) and Sabine Reinfeldt (Chalmers Univ. of Technol., SE-41296 Göteborg, Sweden)

Until recently, own-voice issues for hearing aid users were dominated by the occlusion effect. However, with the so-called open fittings, the occlusion-related own-voice problems have been significantly reduced. There are groups of hearing-aid users, though, for whom open fittings are unavailable, either because of high-frequency gain requirements or because of ear geometry. Of even bigger importance is the fact that hearing-aid users still have own voice issues, even when their occlusion problems have essentially been solved. A model is presented for perception of the own voice comprising an estimation of the own voice transmission by airborne and structure borne sound. Parameters of the model are the occlusion effect from the ear mould and ventilation tubes, gain settings of the hearing aid, and transmission properties of the middle ear during vocalization. The results show that both the airborne and structure borne sound transmission depend on both amplifier and acoustical parameters. A ventilation tube in the ear-mould may introduce a resonance that causes extreme alterations of the own voice spectral content. Although the model is far from complete, it can provide useful insights on own-voice related issues and give design directions to alleviate, at least partly, own-voice problems related to hearing aid usage.

11:45

3aEA11. Transverse variations of sound pressure in an ear canal occluded by a hearing aid. Michael R. Stinson and Gilles A. Daigle (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON, Canada K1A 0R6)

The sound field in a model ear canal has been investigated experimentally and theoretically. The canal was cylindrical, 7.5-mm diameter, and terminated with a Zwislocki coupler to represent absorption by the human middle ear. The outer end of the canal was driven by a test hearing aid, with the acoustic output from a receiver entering the canal through a 1-mm port. The hearing aid was provided with a 20-mm-long vent, either 1 or 2 mm in diameter. The sound field inside the canal was measured using a specially designed 0.2-mm-diam probe microphone [Daigle and Stinson, J. Acoust. Soc. Am. **116**, 2618 (2004)]. In parallel, calculations of the interior sound field were performed using a boundary element technique. Both calculations and experiment reveal large transverse variations of sound pressure level, as much as 25 dB at 8 kHz, near the inner face of the hearing aid. Variations are greatest near the outlet port of the receiver and the vent hole. Deeper into the canal, the transverse variations are less significant and, at depths greater than 5 mm, only a longitudinal variation remains. [Work supported by Gennum Corp.]

3a THU. AM

Session 3aNS

Noise, Speech Communication, and ASA Committee on Standards: Progress in Hearing Protection and Hearing Conservation I

Ellliott H. Berger, Cochair
E-A-R Aearo Co., 7911 Zionsville Rd., Indianapolis, IN 46268-1657

Hiroaki Takinami, Cochair
RION Co., Ltd., 3-20-41 Higashimotomachi, Kokubunji, Tokyo 185-8533, Japan

Chair's Introduction—7:35

Invited Papers

7:35

3aNS1. Results from the National Institute for Occupational Safety and Health Administration/Environmental Protection Agency (NIOSH/EPA) interlaboratory comparison of American National Standards Institute (ANSI) S12.6-1997 Methods A and B. William J. Murphy (Hearing Loss Prevention Team, NIOSH, MS C-27, 4676 Columbia Pkwy., Cincinnati, OH 45226-1998, wjm4@cdc.gov), Elliott Berger (Aearo/E-A-R Corp., Indianapolis, IN 46268-1657), Dan Gauger (Bose Corp. Framingham MA 01701-9168), Brad Witt (Howard Leight Industries, San Diego, CA 92154), Rich McKinley (U.S. Air Force Res. Labs., Dayton, OH 45433), Samir Gerges (Laboratório de Vibraçöes e Acústica, Florianópolis, SC, Brazil), and William Ahroon (U.S. Army Aeromedical Res. Lab., Fort Rucker, AL 36362-0577)

The Environmental Protection Agency and the National Institute for Occupational Safety and Health sponsored the completion of an interlaboratory study (E-A-RCal[SM] Laboratory, Howard Leight Industries, Brazil Acoustics and Vibration Laboratory, NIOSH Robert Taft Laboratories, US Air Force Research Laboratory, and US Army Aeromedical Research Laboratory) to compare two fitting protocols described in ANSI S12.6-1997 (R2002). Method A (informed user fit) and Method B (naive subject fit) were implemented according to the standard. Six models of hearing protection were tested in the study: two earmuffs, foam, premolded, and custom-molded earplugs, and canal caps. The muffs were counterbalanced both in testing order and the occluded and unoccluded conditions. The earplugs were counterbalanced across testing order and occlusion. Each laboratory tested 24 subjects for both methods and each protector. This paper reports on the results from this study. Particularly, the comparison of the statistical errors for each method and protector will be evaluated. In general the lower, 80th percentile, estimates of the expected protection were quite similar. The upper, 20th percentile, tended to exhibit effects that were dependent upon the method and the laboratory. [Work supported by EPA interagency agreement DW75921973-01-0.]

7:55

3aNS2. Methods of measuring the insertion loss of active noise reduction devices. Dan Gauger (Bose Corp., M.S. 271E, 145 Pennsylvania Ave., Framingham, MA 01701, dan_gauger@bose.com), Elliott Berger, Ronald Kieper (E-A-R/Aearo Technologies, Indianapolis IN 46268), John Casali, and Chuck Perala (Virginia Tech, Blacksburg VA 24061)

As active noise reduction (ANR) becomes more common, the need has developed for the measurement of device attenuation, passive and active combined, for use in noise protection calculations. An experiment was conducted involving both REAT and MIRE measurements on five circumaural and two supra-aural ANR devices in two different laboratories. Subject selection and instruction were consistent with the ANSI S12.6-1997 Method B fitting paradigm for all measurements. MIRE was measured both using microphones mounted on canal-occluding plugs near the entrance of the canal ("shallow MIRE") and using concha-mounted microphones with probe tubes extending near to the tympanic membrane and an unoccluded canal ("deep MIRE"); the latter measurements were done with the same device fitting trials as the REAT measurements. The two MIRE measurements were quite similar, with the shallow MIRE values typically better by margins of only 0.5 to 1.0 dB at all frequencies. A positive REAT minus MIRE difference at the lower frequencies consistent with the effects of physiological noise masking was observed. Various analyses of the data are presented and a recommendation is made for how ANR device total attenuation might be measured in a future standard as part of a hearing protector rating regime encompassing both passive and active devices.

8:15

3aNS3. Variability of field-based hearing protection device attenuation measurements. Richard Neitzel, Sebrina Somers, and Noah Seixas (Univ. of Washington, DEOHS, 4225 Roosevelt Way, NE #100, Seattle, WA 98105)

HPD attenuation is highly dependent on individual-specific fit. Prediction of an individual's attenuation depends on measurement system accuracy and attenuation variability due to HPD refitting. This study compared attenuation measurements from two systems [one real-ear-attenuation-at-threshold (REAT) and the other microphone-in-real-ear (MIRE)] on workers using two earplugs (foam and custom-molded). Workers' perceptions of the earplugs were also evaluated. Individual's attenuation results were summarized as personal attenuation ratings (PARs). Variability in PARs from between-subject, between-day, and within-day differences was used to

determine the lower confidence limit, or uncertainty factor (UF), of an individual's attenuation. The custom-molded earplug PARs achieved a higher percentage of labeled attenuation than did the foam earplug with both test systems, and also had higher overall acceptance among workers. MIRE PAR levels were lower than REAT levels for both earplugs, but the relationship between the two test systems was highly variable. The MIRE system had lower within-day variability than the REAT system. One individual's MIRE results were highly influential; removal of these results greatly reduced the UF for the custom-molded earplug/MIRE combination. UFs ranged from 8.8 to 13.5 dB. These findings highlight the importance of evaluating variability in individuals' HPD attenuation, rather than relying on single measurements.

8:35

3aNS4. Has the Occupational Safety and Health Administration record-keeping rule changed hearing loss prevention programs? Laurie Wells (Assoc. in Acoust., Inc., 2283 Danielle Court, Loveland, CO 80537, lwellshear@cs.com)

The Occupational Safety and Health Administration (OSHA) record-keeping rule 29 CFR part 1904.10, changed the definition of recordable hearing loss and created a separate column on the OSHA 300 form to record noise-induced hearing loss (NIHL). In an effort to identify possible effects of the record-keeping rule change, an informal written survey was distributed at the National Hearing Conservation Association (NHCA) annual conference in February, 2006. NHCA conference attendees, because they specialize in hearing loss prevention, are likely to be familiar with hearing conservation issues so their perceptions are of particular interest. A total of 36 attendees, including audiologists, audiometric technicians (OHC), industrial hygienists (IH), mobile test company owners, occupational medicine physicians (OMP), and occupational health nurses (OHN) completed the ten-question survey. In summary, there appears to be a consensus that the everyday practice of hearing loss prevention has been affected by the record-keeping rule change, with both positive and negative influences. Employers seem to be paying more attention to HLPPs, which is, by itself, a positive outcome. Somewhat alarming however, is the finding suggesting that recordable STSs may be being ''lined-out'' or ''denied'' inappropriately in order to keep recordable rates artificially low.

8:55

3aNS5. Hearing effects of nonoccupational sound exposure evaluated by temporary threshold shift. Hiroaki Takinami (RION Co., Ltd., 3-20-41 Higashimotomachi, Kokubunji Tokyo 185-8533 Japan, hiro@rion.co.jp) and Shoji Aono (Osaka Univ., 1-2 Yamadaoka, Suita, Osaka 565-0871, Japan)

In the 1960s, the relation between the acoustical characteristics of sound exposure and the sound-induced temporal threshold shift (TTS) of people were carefully investigated by Shoji $et\ al.$ [J. Acoust Soc. Am. **48** 978−987 (1970); 6th ICAA−2−8, pp. A9−A12 (1968)]. Based on these studies, an equation that estimates TTS at various test frequencies was proposed as a function of the TTS−critical−band levels. In this presentation, therelation between the sound exposure and the TTS will be reviewed, and a prototype TTS meter that estimates TTS in real time operation will be introduced. The hearing effects of some nonoccupational sounds at different places (Pachinkoparlor, disco, and livehouse) were evaluated with the TTS levels as the indicator. The sounds from portable musicplayers and music instruments of or chestra were also evaluated. The TTS levels were measured for shorttime sound exposure and compared to long−term estimates using the TTS meter. In some cases, after 2− or 4−h sound exposure, the estimated TTS levels at the test frequency of 4 kHz reached almost 20 dB.

9:15

3aNS6. Japanese young people's recreational sound environment and hearing protection. Yôiti Suzuki and Fumitaka Saito (Res. Inst. of Electr. Commun., Tohoku Univ., Katahira 2-1-1, Aobaku, Sendai 980-8577, Japan)

Intense sound may damage hearing even if the listeners are willing to listen to it. Risk of hearing disorders due to leisure time sound is an important problem in hearing conservation. A questionnaire survey of Japanese young people's leisure time activities showed that in Japan concerts rather than discos seemed more influential and in comparison with the results in Finland, the awareness of the danger of intense sound was much less. The sound levels while listening to music with loudspeakers and at pop-music concerts were measured. While listening levels with loudspeakers were quite modest (avg.: $L_{Aeq}=65$ dB for favorite music pieces), at the two concerts the L_{Aeq}'s at audience positions ranged from 95 to 105 dB. Since the pop music performed at the concerts was mild relative to typical values in the present Japanese music scenes, this level may be regarded as near the minimum that could be expected for such activities. Additionally, L_{Aeq} while playing one of ten kinds of musical instruments at the player's ear was measured. The results ranged from 76 (viola) to 110 dB (trombone). Consideration of hearing protection at music scenes and education regarding its value seem very important.

Contributed Papers

9:35

3aNS7. Variability in real-ear attenuation at threshold evaluations of hearing protection devices. William A. Ahroon, Melinda E. Hill, and Kimberly B. Lukas (U.S. Army Aeromedical Res. Lab., Fort Rucker, AL 36362)

The American National Standard Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors (ANSI S12.6-1997 [R2002]) contains explicit requirements for the physical parameters of the test facility used in measuring the performance of hearing protection devices and in subject selection and device fitting. However, the standard curiously does not provide a similar level of detail when considering the psychophysical procedure used to define ''threshold'' other than when an automatic recording audiometer is used. Most facilities conducting ANSI-based measurements of hearing protector performance appear to use a Bekesy tracking procedure, but other methods such as a modified method of limits or a method of adjustment procedure are implemented as well by some facilities. Following several tests of hearing protection over the past year, several examples of large variability caused by very low attenuation results in one or both runs in a few subjects have been noted. The data from these unusual subjects are analyzed and presented along with recommendations for the psychophysical procedures to be included in future versions of the standard.

3aNS8. Comparison of passive versus active noise reduction headsets in light aircraft cockpit noise: Communications intelligibility and workload. John Casali, Brian Valimont, Jeffrey Lancaster (Auditory Systems Lab., Virginia Tech, 250 Durham Hall, Blacksburg, VA 24061, jcasali@vt.edu), and Dan Gauger (Bose Corp., Framingham, MA, 01701)

Under instrument flight rules (IFR) conditions in Cessna 172R (95 dBA) cockpit noise, eight IFR-licensed pilots flew four 3.5-h nonstop flights in an FAA-certified, single-engine flight simulator, while wearing one of four communications headsets in each flight. All flights imposed an equivalently high level of air traffic control communications, at three Speech Transmission Index (STI) values: 0.30, 0.50, and 0.70. Perceptual and psychomotor workload were also controlled. In addition to a battery of flight-related metrics (some reported elsewhere), an objective, flight-embedded measure of communications intelligibility was obtained—the number of readbacks required of the pilot to attain 100% intelligibility on each command. An ANOVA revealed that as a group, the ANR headsets (Sennheiser HMEC300, Bose Aviation X, LightSPEED Thirty3G) required significantly fewer readbacks (averaging 1.2–1.3) than did the passive David Clark H-10-13.4 (averaging 1.5). Flight control performance, as reflected by deviations in assigned magnetic heading, altitude, and airspeed, was highest during flights with the Bose headset. Pilots' workload estimates were measured using the modified Cooper-Harper scale, with 1 representing lowest workload and 9 representing highest. The three ANR headsets were associated with average workload ratings of 3.4–3.5, while the passive headset yielded a higher average workload of 4.4.

10:05–10:15   Break

3aNS9. An evaluation of the acoustic properties of an improved combat arms earplug. Mary S. Binseel and Tomasz R. Letowski (U.S. Army Res. Lab., AMSRD-ARL-HR-SD, Aberdeen Proving Ground, MD 21005, mary@arl.army.mil)

This study investigates the performance of an improved version of the combat arms earplug (CAE). U.S. Army soldiers must operate in environments which can go from quiet to noisy in a moment. They also operate in environments where the predominant noise hazard can be either continuous or impulse noise. The current earplug of choice to address these varying protection needs is the CAE. The CAE is a two-sided earplug which provides linear attenuation for steady-state noise exposure on one side and nonlinear protection for impulse noise protection on the other. Although feedback from the field has been favorable regarding the functioning of the CAE, other concerns such as fit, comfort, size, and the need to remove, turn, and reinsert the earplug to change the modality have been raised. In order to address these concerns, the Army Research Laboratory provided funding for the development of an improved CAE. The improved CAE was evaluated for attenuation of continuous and impulse noise, soldier performance in localization and speech intelligibility tasks, and soldier acceptance on issues such as ease of use, perceived protection, equipment compatibility, sizing, and comfort. Results of these experiments will be discussed.

3aNS10. Design of linear time-domain filters for hearing protector modeling. Jeffrey S. Vipperman (Dept. of Mech. Eng., Univ. of Pittsburgh, 648 Benedum Hall, Pittsburgh, PA 15261) and William J. Murphy (Natl. Inst. for Occupational Safety and Health, Cincinnati, OH 45226-1998)

Accurately modeling hearing protectors for predicting the hazard underneath them is of interest to many. This project modeled the single transfer path through the hearing protector from the free field to the ear drum, as an extension to the frequency-domain modeling method presented by Price and Kalb [J. Acoust. Soc. Am. **103**, 2878 (1998)]. First, the real ear attenuation at threshold (REAT) data from the hearing protectors (muff and plugs) were converted to linear quantities and curve fit to form

a continuous function of frequency, while preserving the low-pass characteristic. The resulting magnitude was combined with the head-related transfer function (HRTF) to form the magnitude of the total transfer function. Using complex cepstrum, a stable, complex frequency response function (FRF) was created by assuming a minimum phase transfer path. System identification methods were used to fit the FRF data to a parametric model that was converted to an infinite impulse response (IIR) digital filter. Subject-fit REAT data were used in order to represent real-world usage. Imperfect usage was modeled by derating the subject fit data by 1 standard deviation (SD). Tests on an artificial head are conducted to evaluate the filters in the time and frequency domains. [This research was supported wholly by NIOSH.]

3aNS11. A source of high-level acoustic impulses for the measurement of hearing protectors. Jan Zera and Rafal Mlynski (Central Inst. for Labour Protection, Natl. Res. Inst., Czerniakowska 16, 00-701 Warsaw, Poland)

The paper describes a laboratory source for generating high-level, A-type acoustic impulses. The basic part of the source is a steel cylinder closed at one end with elastic foil. The blast is produced by an abrupt break of the foil caused by an increased pressure of air compressed inside the cylinder. By adjusting the air pressure inside the cylinder the impulses peak SPL may be set within a range of 150–170 dB. As the peak SPL is increased from 150 to 170 dB, the rise time and A duration of impulses decrease respectively from 350 to 14 s and from 1.1 to 0.4 ms. An advantage of the source is that it produces impulses with highly repeatable parameters; for a given value of air pressure inside the cylinder, standard deviation of peak SPL distribution is only 1.3 dB. The source is omnidirectional in the horizontal plane and has a nearly circular directional pattern in the vertical plane. The source is a good alternative for generating impulses with the use of explosives and is particularly useful for measurements made indoors, in the laboratory. [Work supported by the Polish Ministry of Science and Education, Grants T07B00428 and SP03.1].

3aNS12. Applicability of impulse noise hearing damage risk criteria (DRCs) to hearing protectors. Jan Zera and Rafal Mlynski (Central Inst. for Labour Protection, Natl. Res. Inst., Czerniakowska 16, 00-701 Warsaw, Poland)

Hearing damage risk criteria, DRC, concerning exposure to impulse noise, developed by CHABA [Ward CHABA Report (1968)], Pfander [J. Acoust. Soc. Am. **67**, 628–633 (1980)], Smoorenburg [*New Perspectives on Noise*] (Raven, New York, 1982), pp. 471–490], and Dancer [Acust. Acta Acust. **3**, 539–547 (1995)], have been destined for impulses encountered in open space, in free-field conditions. To adapt these criteria for conditions in which hearing protective devices (HPDs) are used, a standard procedure was proposed in the literature (e.g., by Pfander) that consists in subtracting the attenuation of the HPD, typically 25 dB, from the impulse's peak SPL. Such a simple method does not take into account the change of the signal waveform under the HPD, usually resulting in increased A, C, or D duration. In this study, the subtraction method was verified by a comparison of impulse noise DRC with the parameters of acoustic impulses measured under the earmuff: peak SPL, rise time, and A, C, and D duration. Results show that the subtraction method underestimates the possible hearing damage risk posed by exposure to impulse noise when a hearing protector is used. [Work supported by the Polish Ministry of Science and Education, Grants T07B00428 and SP03.1].

3aNS13. Performance of sound restoration hearing protection in impulsive noise. William J. Murphy (Hearing Loss Prevention Team, Natl. Inst. for Occupational Safety and Health, 4676 Columbia Pkwy., MS C-27, Cincinnati OH 45226-1998)

Hearing protection devices (HPDs) are the principal means of protecting the hearing of a person against harmful levels of noise in highly impulsive noise environments. New HPDs which utilize sound restoration circuitry were measured using a mannequin to assess the performance of

these protectors in response to small-arms weapons fire. The results for the peak reduction and attenuation were analyzed for both indoor and outdoor measurements. The performance of the protectors exhibited little dependence with level for impulses between 150 and 170 decibels. The passive performance for a single protector characterized the active performance for these high-intensity impulses. The results were also evaluated using the US Army AHAAH cochlear model. As the peak sound-pressure level underneath the different models of protectors increased, the estimated risk of hearing loss increased. [Work supported by EPA Interagency Agreement DW75921973-01-0.]

11:30

3aNS14. Criteria levels for nonoccupational noise exposure. Leslie Blomberg (Noise Pollution Clearinghouse, Box 1137, Montpelier, VT 05601)

Selecting criteria levels for nonoccupational noise exposure requires consideration of several factors. These include: (1) the need to protect public health and safety; (2) the diverse sources of nonoccupational noise exposure; (3) the inconsistent patterns of exposure, for example, the potential for multiple exposures in 1 day, or none; (4) complications caused by workplace exposure; (5) the fact that those exposed include children and minors; (6) the inability of the general public, without training or equipment, to understand traditional means of expressing noise exposure; (7) difficulties educating the public about nonoccupational noise; and (8) complications that may arise if levels are mandated or adopted by government bodies. Given these considerations, the paper argues there are actu-ally three criteria levels needed: (1) one that delimits safe noise exposure; (2) one that reflects increased levels of risk and the need for increased vigilance in protecting hearing; and (3) one that indicates unsafe exposure, for which hearing conservation should be practiced. Moreover, new and simpler ways of determining these levels are needed. Specific levels are recommended, and a method developed by the Noise Pollution Clearing-house for use by the general public to determine if a sound is too loud is presented.

11:45

3aNS15. Dosimeter measures of sound exposure experienced by university students. William A. Yost, Noorah Koita, Robert Maslo, Prasad Patel, Ruchita Patel, and Brendan Ringhouse (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

There is increased concern about excessive sound exposure, especially in students, resulting from the use of Ipods and similar devices. Such exposure needs to be considered in light of the daily noise dose such users experience. In this study five university students wore noise dosimeters for 10 h per day for 5 days, and noise levels were recorded every minute. The students also maintained a diary of their activities (activities were noted at least every half hour) during the time they wore the noise dosimeters. The diary was standardized so that each student used the same diary format. Results indicating the noise levels, noise dose, and the relationship of the levels to the students activities will be presented and discussed. In general the noise exposure experienced by these students was usually mild (less than 60 dBA) to moderate (between 60 and 75 dBA).

Contributed Poster Paper

Poster paper 3aNS16 will be on display from 7:30 a.m. to 12:00 noon. The author will be at the poster from 12:00 noon to 12:15 p.m.

3aNS16. Review of literature on hearing damage by personal stereo. Dorte Hammershøi and Rodrigo Ordoñez (Dept. of Acoust., Aalborg Univ., Fredrik Bajers Vej 7-B5, DK-9220 Aalborg Ø, Denmark, dh@acoustics.aau.dk)

In the 1980s and 1990s there was a general concern for the high levels that personal stereo systems were capable of producing. At that time no standardized method for the determination of exposure levels existed, which could have contributed to overly conservative conclusions. With the publication of ISO 11904-1:2002 and 11904-2:2004, previous studies can be viewed in a different light, and the results point, in our opinion, at levels and listening habits that are of hazard to the hearing. The present paper will review previous studies that may shed light over the levels and habits of contemporary personal stereo systems, which can produce even higher levels and are of even wider use.

Session 3aPA

Physical Acoustics: Cavitation I

Claus-Dieter Ohl, Cochair

Univ. of Twente, Faculty of Science and Technology, P.O. Box 217, 7500 AE Enschede, The Netherlands

Shigeo Hayashi, Cochair

Univ. of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

Chair's Introduction—8:30

Invited Papers

8:35

3aPA1. Attempts to observe bubble fusion in deuterated acetone.　Carlos G. Camara, Seth J. Putterman, Brian Naranjo　(Phys. Dept., UCLA, 2-202 Knudsen Hall, Los Angeles, CA, 90095, camara@physics.ucla.edu), Stephen D. Hopkins　(Intel Corp. Portland, OR, 97205), and Kenneth S. Suslick　(Univ. of Illinois, Urbana, IL 61801)

A pulsed neutron generator (PNG) and a radioactive source have been used to seed cavitation at dynamic pressures ranging up to 35 atm. No fusion signal above background has been observed. Our upper bound is less than $1E-4$ of the signal claimed by R. P. Taleyarkhan *et al.* [Phys. Rev. Lett. **96**, (2006)]. Reasons for the failure of bubble fusion in acetone and future directions will be discussed. [Research funded by DARPA. We thank Brian Kappus for valuable discussions.]

8:55

3aPA2. Influence of ultrasonic frequency on sonochemical reactions.　Kyuichi Yasui, Toru Tuziuti, Teruyuki Kozuka, Atsuya Towata, and Yasuo Iida　(Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), Moriyama-ku, Nagoya 463-8560, Japan, k.yasui@aist.go.jp)

Numerical simulations of the pulsation of an air bubble have been performed for various ultrasonic frequencies (20 kHz to 1 MHz) and pressure amplitude taking into account the effect of nonequilibrium chemical reactions inside a bubble and that of nonequilibrium evaporation and condensation of water vapor at the bubble wall. The calculated results have revealed that as ultrasonic frequency decreases, the mole fraction of water vapor inside a bubble increases at the end of the bubble collapse. While the main oxidant created inside a vaporous bubble, which is defined as a bubble consisting mostly of water vapor at the end of the bubble collapse, is OH radical, that inside a gaseous bubble, which is defined as a bubble consisting mostly of noncondensable gas (air) at the end of the bubble collapse, is H_2O_2 or O atom when the bubble temperature is less than or more than 7000 K, respectively. Although the optimum acoustic amplitude for the oxidant's creation inside a bubble increases as the ultrasonic frequency increases, the range of the suitable acoustic amplitude for the oxidants creation becomes wider. It possibly results in the optimum ultrasonic frequency for the oxidant's creation reported in the literatures (200–600 kHz).

9:15

3aPA3. Comparison of sonoluminescence and water vapor plasma spectra.　Masanori Sato, Nagaya Okada　(Honda Electron. Co., Ltd., 20 Oyamazuka, Oiwa-cho, Toyohashi, Aichi 441-3193, Japan), Hideo Sugai, Tatsuo Ishijima, and Hirotaka Hotta　(Nagoya Univ., Nagoya, Aichi 464-8603, Japan)

An experimental study was carried out to find the mechanism of OH spectra generation through sonoluminescence. Line spectra of OH, H, and O were observed very clearly in water vapor plasma generated by 2.45-GHz microwave irradiation under the pressure of 40 hPa; however, broadband continua were not detected. Multibubble sonoluminescence (MBSL) and single-bubble sonolumines- cence (SBSL) have continua, and OH spectra were detected occasionally in MBSL. However, OH spectra were not detected in SBSL and higher temperature MBSL. The sources of emission in MBSL and SBSL were considered to be identical and differences in spectra of OH bands were attributable to the higher temperature. Experimental results obtained using low temperature water vapor plasma confirm that OH spectra depend on the bubble's interior plasma temperature: that is, at low temperatures, spectra of OH (308 nm), Ha, Hb, and O are clearly visible in water vapor plasma.

9:35

3aPA4. Molecular dynamics study on the nucleation and growth of cavitation bubbles.　Shu Takagi　(Dept. of Mech. Eng., Univ. of Tokyo, Bunkyo Ku, 7-3-1 Hongo, Tokyo 113-8656, Japan, takagi@mech.t.u-tokyo.ac.jp) and Shinichi Tsuda　(Univ. of Tokyo, Tokyo 113-8656, Japan)

Molecular dynamics simulations have been conducted for the nucleation and growth processes of cavitation bubbles in a decom- pressed liquid. To investigate the effects of noncondensable gas, the comparison between one-component liquid argon and two- component liquid argon, which includes helium gas molecules as impurities, was made. As a result, we observed that a competing coarsening similar to Ostwald ripening occurs in the one-component liquid argon, while coalescence of bubble nuclei frequently

occurs in the two-component case. In spite of the different growth patterns between the two cases, the values of the growth exponent are close to 1/2 in both cases. However, it has been found that the scaling components that constitute the growth exponent, i.e., the scaling exponent for the total radius of all nuclei and that for the number of nuclei, are different between the two cases. These scaling components are discussed in relation to a driving force for bubble nuclei to grow. In addition, the asymptotic behavior of the size distribution function of bubble nuclei is also discussed using these scaling parameters.

9:55–10:15 Break

Contributed Papers

10:15

3aPA5. Frequency effects on intense sonoluminescence and bubble dynamics in sulfuric acid. Shin-ichi Hatanaka and Shigeo Hayashi (Univ. of Electro-Commun., 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585 Japan, hatanaka@pc.uec.ac.jp)

Emission intensity increased remarkably with decreasing frequency for both multibubble sonoluminescence (MBSL) and single-bubble sonoluminescence (SBSL) in a concentrated sulfuric acid solution dissolving xenon gas. The spectrum of MBSL in the sulfuric acid showed the emission bands of a large peak at 388 nm and small peaks at 360 and 418 nm. The increase of MBSL intensity at lower frequency was responsible for the increase of the band peak at 388 nm in addition to the increase in continuous spectrum. The bubble dynamics in the sulfuric acid was different from that in water. The intense SBSL was emitted when the bubble was moving on an ellipselike trajectory. The intense MBSL was emitted even when multiple bubbles formed bubble clusters, which could not emit light in the water case.

10:30

3aPA6. Experimental analysis of blackbody emission from sonoluminescence in sulfuric acid. Stephen D. Hopkins, Carlos G. Camara, and Seth J. Putterman (Phys. Dept., UCLA, 2-202 Knudsen Hall, Los Angeles, CA 90095)

The spectrum of emission from a single xenon bubble in concentrated sulfuric acid driven at 30 kHz is an excellent fit to Planck's law with a surface temperature of 8000 K. The measured flash width and emission radius are also consistent with blackbody emission. In this study the only fitting parameter available is the temperature [Phys. Rev. Lett. **95**, (2005)]. [Research funded by DARPA.]

10:45

3aPA7. Drop tube generates 10-W flashes of sonoluminescence. Brian A. Kappus, Avic Chakravarty,[a] , and Seth J. Putterman (Phys. Dept., UCLA, 1-129 Knudsen Hall, Los Angeles, CA 90095)

Use of a low vapor pressure liquid such as phosphoric acid, with dissolved xenon in a vertically exited tube, generates ~200-ns flashes of sonoluminescence with a peak power of 10 W. We are in the process of characterizing the bubble motion by use of backlighting, stroboscopic, and streak photography. We will also broach the topic of disequilibrium between atom and electron temperatures. [Research funded by DARPA. We thank Carlos Camara and Shahzad Khalid for valuable discussions.]
[a] Deceased.

11:00

3aPA8. Attempts to observe sonoluminescence from a single bubble driven at 10 MHz. Shahzad Khalid, Carlos G. Camara, and Seth J. Putterman (Phys. Dept., UCLA, 2-202 Knudsen Hall, Los Angeles, CA 90095)

A 10-MHz spherical transducer array is used to generate cavitation at its focus with a field of ~100 atm. Sonoluminescence as well as light-scattering measurements will be discussed. [Research funded by DARPA.]

11:15

3aPA9. Unidentified energy conversions in an oscillating bubble. Karel Vokurka (Phys. Dept., Tech. Univ. of Liberec, 46117 Liberec, Czech Republic) and Silvano Buogo (CNR IDAC, 00133 Roma, Italy)

Acoustic radiation of freely oscillating bubbles has been studied experimentally. The oscillating bubbles have been generated by underwater spark discharges. Recorded pressure waves made it possible to determine the acoustic energy radiated by the oscillating bubbles and the potential energy of the bubbles when attaining the first and second maximum volumes. Comparison of the potential and acoustic energies revealed that about 70% of the potential energy of the bubble available at the first maximum volume has been converted into an unidentified energy. A possible candidate for the energy conversion mechanism seems to be internal converging shock wave in the bubble interior. The presence of this wave can be observed indirectly in a radiated pressure wave. The measured energies are compared with energies computed using a suitable model. This comparison shows that the theoretical values agree with experimental data until the bubble attains the first minimum volume. However, at later times discrepancies are observed. [Work has been partly supported (K.V.) by the Czech Ministry of Education as the research project MSM 46747878501.]

11:30

3aPA10. Optical observation of collapses of two bubbles adhered to the quartz wall under ultrasound irradiation with high-speed video camera: Micro-jet arising from two collapsing bubbles. Kenji Yoshida, Shintaro Nakatani, and Yoshiaki Watanabe (Faculty of Eng., Doshisha Univ., 1-3 Tataramiyakotani, Kyotanabe, Kyoto, Japan, etf1103@mail4.doshisha.ac.jp)

The mechanical action of a micro-jet, which arises from the collapse of bubbles, has been widely used for ultrasonic cleaning and sonoporation in medical field. The generation mechanism of a micro-jet induced by the ultrasound irradiation is focused on. In previous theoretical and experimental researches, the generation mechanism is well known when a single bubble is adhered to a rigid wall. However, its mechanism is not understood in detail when multiple bubbles exist on rigid wall. The optical observation of multi-bubble collapse greatly contributes to the development of micro-jet applications. It was observed that the collapses of two bubbles adhered to the quartz wall in an ultrasound field with high speed video camera (HPV-1, SHIMADZU). Maximum recording rate of this high speed video camera is 1 000 000 frames/second. The emitted ultrasound frequency is 27 kHz. From the observation, one of two bubbles was deformed after these bubbles vibrate. The micro-jet was generated to the opposite side against the deformed bubble when the deformed one touches the other. Then, the direction of the micro-jet to the quartz wall was almost horizontal. The number of micro-jets increased every time the distorted bubble touched the other. Finally, five micro-jets to the various directions were observed at the same instant.

Session 3aPP

Psychological and Physiological Acoustics and Musical Acoustics: Perception of Music and Speech: Evidence for Modularity and for Common Processing Mechanisms

Diana Deutsch, Cochair

Univ. of California, San Diego, Dept. of Psychology, La Jolla, CA 92093

Yoshitaka Nakajima, Cochair

Kyushu Univ., Dept. of Acoustic Design, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

Chair's Introduction—8:30

Invited Papers

8:35

3aPP1. Speech perception: Development from general mechanisms to modules. Patricia K. Kuhl (Inst. for Learning & Brain Sci. and Dept. of Speech and Hearing Sci., Univ. of Washington, Seattle, WA 98195)

Are the mechanisms subserving speech perception modular or general process in nature? I will adopt a developmental perspective to address this question. Using data from cross-cultural, cross-species, and developmental experiments, I will argue that, early in infancy, the mechanisms underlying speech perception are domain general and rely on processing that is not exclusive to speech signals. Language evolved to capitalize on general mechanisms of perception and action, and this ensures that all infants, regardless of the culture in which they are born, initially respond to and learn from exposure to language. As experience with language ensues, neural commitment to the properties of the language to which an infant is exposed results in dedicated neural networks and speech perception becomes more specialized. Finally, I will suggest that adult members of the culture help instill language in the child by adjusting the language they use when addressing infants and children (motherese), a speech style that exhibits interesting acoustic properties. Language input to the child enhances its learnability. A model of the developmental process will be described. [Research supported by NSF and NIH.]

8:55

3aPP2. Auditory events in language and music. Yoshitaka Nakajima and Kazuo Ueda (Faculty of Design, Kyushu Univ., 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan, nakajima@design.kyushu-u.ac.jp)

Perceptual organization of both language and music is based on temporal units that can be called auditory events. Speech syllables and musical notes correspond typically to auditory events. The listener categorizes these events in various ways in order to extract information. We are at a preliminary stage of searching for acoustic cues that the auditory system can utilize to determine auditory events. Averaged intensity of a speech signal within a moving Gaussian temporal window of 15 ms gives cues of syllable boundaries, especially when this averaged intensity is compared with the intensity change obtained with a Gaussian temporal window of 200 ms. The change of intensity obtained through a filter of about 600–2200 Hz compared with the intensity change obtained through an almost flat filter also gives cues of syllable boundaries. Thus, intensity cues and spectral cues both work to mark syllables. This holds both in Japanese and in English. Such intensity cues may work when applied to monophonic or homophonic music, but spectral changes do not correspond clearly to musical notes. Thus, the perceptual mechanisms that determine speech syllables and musical notes can be different. [Work supported by JSPS and 21st Century COE, Kyushu University.]

9:15

3aPP3. Music and language in deaf children with cochlear implants. Takayuki Nakata, Sandra E. Trehub, Yukihiko Kanda, and Haruo Takahashi (Dept of Psych., Nagasaki Junshin Catholic Univ., 235 Mitsuyama-machi, Nagasaki-shi, Nagasaki-ken 852-8558, Japan, Dept. of Psych., Univ. of Toronto, Nagasaki Bell Hearing Ctr., Nagasaki Univ. Grad School of Biomed. Sci.)

Congenitally deaf children with cochlear implants successfully perceive speech in quiet environments, and many produce intelligible speech with normal voice quality. Because implants were designed to optimize speech rather than music perception, timing cues are preserved at the expense of pitch cues. Consequently, adult implant users tend to discontinue musical activities that were enjoyable prior to the onset of deafness. By contrast, congenitally deaf children enjoy a variety of musical activities including listening, singing, and playing instruments. Research in our laboratory suggests that musical activities may enhance child implant users' speech perception skills. For example, initiation of music listening at home by children with cochlear implants predicted their success on a word recognition task. Moreover, earlier ages of implantation were positively associated with enhanced motivation for listening to music. In addition, child implant users were able to identify musical themes that accompanied their favorite television programs. Unlike hearing children, however, they could not recognize the music from the melody (i.e., pitch patterning) alone. Analyses of these children's singing revealed rhythmic patterning similar to that of hearing children but pitch patterning that was grossly distorted. In short, rhythm plays a central role in child implant users' musical experience.

3aPP4. Dissociation of music and speech: Evidence from statistical learning in congenital amusia. Isabelle Peretz (Dept. of Psych., Univ. of Montreal, B.O. 6128, Succ. Ctr. ville, Montreal, QC, Canada H3C 3J7, isabelle.peretz@umontreal.ca) and Jenny Saffran (Univ. of Wisconsin—Madison, Madison, WI 53706)

Individuals who suffer from congenital amusia (commonly called tone-deafness) have experienced life-long difficulties in processing of music. The deficit can be traced down to an impairment in fine-grained pitch discrimination [Hyde and Peretz, Psychol. Sci., (2004)]. Their language abilities appear intact. One possible account for this dissociation between music and speech is that amusics lack normal exposure to music because of their lifelong history of musical failures. In other words, amusics might be able to acquire basic musical abilities if given appropriate exposure. To test this possibility, a group of 11 adults with congenital amusia, and their matched controls, were exposed to a continuous stream of syllables or tones for at least 21 min. Their task was to try to identify words or motifs, defined by transitional probabilities, following the procedure used by Saffran and collaborators [Cognition, (1999) (2001)]. The results show that amusics can learn novel words as easily as controls whereas they systematically fail on musical materials. Thus, limited exposure cannot fully account for the musical disorder. Implications for the domain specificity/generality debate of statistical learning will be discussed.

3aPP5. Nonlinguistic rhythm perception depends on culture and reflects the rhythms of speech: Evidence from English and Japanese. Aniruddh D. Patel, John R. Iversen (The Neurosci. Inst., 10640 John Jay Hopkins Dr., San Diego, CA 92121), and Kengo Ohgushi (Kyoto City Univ. of Arts, Kyoto, Japan)

Does a listener's native language influence the perception of nonverbal rhythms? One way to address this question is to examine the perception of rhythmic grouping, as suggested by Jakobson, Fant, and Halle in *Preliminaries to Speech Analysis*. Grouping refers to the perceptual clustering of events into higher level units such as phrases. Several principles of grouping of auditory patterns, established a century ago, have come to be thought of as innate properties of perception. The supporting data have come entirely from Western cultures, however. In this study, native English and Japanese speakers were tested for their perception of grouping in simple rhythmic sequences of tones. Members of the two cultures showed different patterns of perceptual grouping, demonstrating that basic auditory processes are shaped by culture. Furthermore, measurements of the temporal structure of English versus Japanese speech and music suggest that the observed differences reflect the rhythms of the two languages, rather than rhythmic differences between American and Japanese music. While it has long been known that a listener's native language influences the perception of speech, these results suggest that native language can also exert an influence on general auditory perception at a basic level. [Supported by Neurosciences Research Foundation and by JSPS.]

3aPP6. Two issues at the interface between speech and music. Diana Deutsch (Dept. of Psych., Univ. of California, San Diego, La Jolla, CA 92093) and Trevor Henthorn (Univ. of California, San Diego, La Jolla, CA 92093)

In what ways are speech and music subserved by domain-specific mechanisms, and in what ways by common neural substrates? This issue is addressed with respect to absolute pitch, which has generally been considered to be a musical faculty accessible only to rare individuals. Based on several findings obtained from speakers of tone languages, it is argued that absolute pitch originally evolved to subserve speech, and that it may be acquired readily during the first year of life, if infants are given the opportunity to associate pitches with verbal labels. It is further conjectured, based on other findings, that the early acquisition of pitch accent languages, such as the Tokyo and Osaka dialects of Japanese and the Hamkyeng and Kyengsang dialects of Korean, might also be conducive to the acquisition of absolute pitch. Perceptual relationships between speech and song are next considered. These generally differ in their features; however, a spoken phrase can be perceptually transformed to appear as song rather than speech, simply by repeating it several times over without any alteration in the signal (Deutsch, 2003). This phenomenon is demonstrated, its parameters are explored, and its implications for the mechanisms underlying perception of speech and song are discussed.

Contributed Papers

3aPP7. Training musicians and nonmusicians to discriminate Mandarin tones. Terry L. Gottfried and Grace Y.-H. Ouyang (Dept. of Psych., Lawrence Univ., Appleton, WI 54912-0599, rewgottt@lawrence.edu)

Musicians and nonmusicians were compared in their ability to learn Mandarin tone contrasts. Participants were pretested to assess their accuracy in tone discrimination. Some were then given six sessions of articulatory and acoustic training over 3 weeks; others were given no training. At the end of 3 weeks, participants were again tested on their discrimination accuracy. Listeners judged whether the tones (tone 1, high level; tone 2, mid rising; tone 3, low dipping; tone 4, high falling) of two different Mandarin syllables were the same or different. In some trials, listeners discriminated tones in words spoken by two speakers; in other trials, they discriminated tones in words spoken by one speaker only. The pretest revealed that musicians are significantly more accurate in discriminating Mandarin tones than nonmusicians. Musicians and nonmusicians who received the training regimen performed more accurately than untrained listeners on the posttest discrimination. Musicians showed significantly less improvement than nonmusicians with training, probably due to ceiling effects (their posttest accuracy was over 95%). Two-speaker discrimination trials were significantly more difficult for all listeners, but especially for nonmusicians. Discrimination of tones 2 and 3 was most difficult for all listeners, but showed considerable improvement after training.

3aPP8. The effects of musical experience on linguistic pitch perception: A comparison of Norwegian professional singers and instrumentalists. Maren Helene Rø, Dawn Behne (Dept. of Psych., Norwegian Univ. of Sci. and Technol., N-7491 Trondheim, marenhel@stud.ntnu.no), and Yue Wang (Simon Fraser Univ.)

Speech prosody and music share tonal attributes well suited for studying cross-domain transfer effects. The present study investigates whether the specific pitch experience acquired by professional singers and instrumentalists transfers to the perception of corresponding prosodic elements in a native and non-native language. Norwegian and Mandarin words with tonal distinctions, together with corresponding hummed tones, were presented dichotically in a forced attention listening test to three groups of native Norwegian listeners: professional singers, professional instrumentalists, and nonmusicians. While instrumentalists and singers were both more accurate (higher percent correct for both ears) than nonmusicians for Mandarin linguistic and hummed tones, only instrumentalists showed positive transfer to corresponding native Norwegian stimuli. Results indicate a pattern of perceiving tonal distinctions that mirrors the pitch experience acquired through professional vocal and instrumental training: Instrumentalists generally appear to rely on an autonomous, categorical pitch strategy, consistent with their primary training with discrete, fixed intervals. Singers, on the other hand, tend to use a similar strategy for nonlinguistic pitch perception, but not in processing native tonal contrasts. Overall, pitch experience acquired through musical training appears to have a positive transfer to non-native speech perception.

3aPP9. Effects of musical experience on linguistic pitch perception training. Dawn Behne (Psych., Norwegian Univ. of Sci. and Technol., 7491 Trondheim, Norway), Yue Wang (Simon Fraser Univ., Burnaby, BC V5A 1S6, Canada), Maren Helene Rø, Ann-Karin Hoff, Hans Andreas Knutsen, and Marianne Schmidt (Norwegian Univ. of Sci. and Technol., 7491 Trondheim, Norway)

Research continues to address the extent to which music ability transfers to other tasks and processes, among them speech perception. The current study examines the transfer of music experience to native and non-native linguistic pitch perception and tracks this process during training. Participants were nonmusicians (NMs) and music conservatory students (CMs), all of whom were native Norwegian listeners. Participants were tested with Norwegian and Mandarin materials in a tone-based and intonation-based directed attention dichotic listening task, administered five times during 2 weeks. Results show that training generally leads to increased accuracy in linguistic pitch perception. CMs are more accurate than NM in both tasks, and maintain this advantage throughout training. For both groups, training led to a decreased difference in performance between ears, with notably different patterns of ear improvement between NMs and CMs. Across languages, whereas CMs had improved accuracy in both tasks, NM showed no improvement in the tone task for either language. Findings indicate that musically trained listeners' experience positively transfers to native and non-native linguistic pitch perception, an advantage which, in training, affords them more extensive improvement across linguistic pitch tasks and in particular for a non-native pitch task where the native linguistic system does not interfere.

11:35–11:55
Panel Discussion

THURSDAY MORNING, 30 NOVEMBER 2006　　　　　MOLOKAI ROOM, 8:00 A.M. TO 12:00 NOON

Session 3aSC

Speech Communication and Psychological and Physiological Acoustics: Second Language Acquisition II (Poster Session)

Sang Yee Cheon, Cochair
Univ. of Hawaii, Dept. of East Asian Languages and Literature, Honolulu, HI 96822

Nobuaki Minematsu, Cochair
Univ. of Tokyo, Dept. of Frontier Informatics, 5-1-5 Kashiwanowa, Kashiwa-shi, Chiba 277-8562, Japan

Contributed Papers

All posters will be on display from 8:00 a.m. and 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. and 12:00 noon.

3aSC1. A computer-assisted instruction system for self-teaching of discriminating Chinese tones based on their acoustical properties. Shizuo Hiki, Kazuko Sunaoka, and Song Liu (Waseda Univ., 2-579-15 Mikajima, Tokorozawa, 359-1192 Japan, hiki@waseda.jp)

A computer-assisted instruction (CAI) system for self-teaching to discriminate Chinese tones was developed, and provided through the internet to Japanese college students of Chinese language class. The data of acoustical analysis and perceptual judgment on the Chinese tones played an important role in designing the system. The result of observation on characteristic errors in discriminating Chinese tones by the Japanese students was also utilized. This system offered three kinds of use for self-teaching:

testing, training, and reviewing. The original word lists consisted of 15 bisyllabic words with every combination of the four tones of Standard Chinese, involving different rates of unvoiced consonants, which disturbed the tone discrimination. In the partial word lists, only words having tone 2 and/or tone 3, which were difficult for the beginners, were excerpted from the original lists. Their speech samples with both female and male voices were accommodated. Temporal changes in fundamental frequency were extracted from the speech samples and displayed as visual cues in the training. It was also possible in the training to present only the words answered incorrectly in the preceding testing. By selecting suitable uses and word lists with the progress of self-teaching, the required time for achieving the goal was reduced.

3aSC2. Providing new methods for teaching Japanese speech communication for learners of Japanese as L2: Developing teaching materials using MRI-movies and natural speech data. Ryoko Hayashi, Chunyue Zhu, Toshiyuki Sadanobu, Jumpei Kaneda (Grad. School of Cultural Studies and Human Sci., Kobe Univ., Nada-ku Tsurukabuto 1-2-1, Kobe 657-8501, Japan, rhayashi@kobe-u.ac.jp), Donna Erickson (Showa Music Univ., Kanagawa 243-8521, Japan), Nick Campbell (ATR, Hikaridai 2-2-2, 619-0288, Japan), and Miyoko Sugito (Inst. for Speech Commun. Res., Nara 631-0014, Japan)

It is critical to provide people learning Japanese as a second language information about natural speech, since there is wide variability of articulation and speaking styles associated with various social attitudes and/or expressions in Japanese. For developing teaching materials for English and Chinese learners of Japanese as a second language, two new methods are demonstrated. One utilizes MRI-movies that dynamically demonstrate differences among vowel articulations in Japanese, English, and Chinese. This approach is effective for teaching good pronunciation, especially with regard to consonant cluster production, since the timing of the articulation for consonant clusters is visibly presented. The other is audio-visual data of natural speech in Japanese that demonstrate several typical expressions, e.g., wry face and strained (laryngealized) voice for asking favors politely. This type of material shows not only variations of speech communication in Japanese but also cultural differences among native speakers of Japanese, English, and Chinese. [Work supported by MEXT.]

3aSC3. Effects of training with sentences at varied speaking rates on native English speakers' perception of Japanese vowel length. Yukari Hirata, Emily Cullings, Elizabeth Whitehurst, Jacob Whiton (Dept. of East Asian Lang. and Lit., Colgate Univ., 13 Oak Dr., Hamilton, NY 13346), and Stephen Lambacher, G. (Univ. of Aizu, Aizu-Wakamatsu, Fukushima 965-8580, Japan)

In training non-native speakers to perceive difficult second language (L2) phonetic contrasts, providing variation in speakers and phonetic contexts helps the formation of L2 categories [Pisoni and Lively, in *Speech Perception and Linguistic Experience* (1995), pp. 433–459. The present study tested the hypothesis that high stimulus variability in speaking rate would have positive effects for L2 learners. Native English speakers' ability to identify Japanese vowel length was examined with three types of training: slow-only, fast-only, and slow-fast. Their perceptual abilities were measured with a pretest and posttest and compared with a control group that was not trained. Test stimuli were sentences containing Japanese disyllables spoken at three rates (slow, normal, fast). Participants were trained to identify the length of the second vowel of disyllables embedded in a sentence and they received immediate feedback. The pretest to posttest improvement of the slow-only (8.2%) and slow-fast (8.7%) training groups was significantly greater than the control group's improvement (3.9%). However, the fast-only (7.7%) training group's improvement did not significantly differ from the control group's improvement. In addition, slow-fast training, unlike other training, had an advantage in performance on the fast rate stimuli. Implications for the variability theory will be discussed. [Work supported by NSF Grant No. BCS0418246.]

3aSC4. Factors affecting the success of non-native vowel perception training. Kanae Nishi and Diane Kewley-Port (Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

Previously, Nishi and Kewley-Port [J. Acoust. Soc. Am. **118**, 1900 (2005)] reported the successful outcomes of an intensive vowel perception training protocol for Japanese and Korean learners of English. Their results indicated that vowel training worked best when large numbers of vowels in nonsense words were presented within the training blocks. In this report, additional analyses were performed in order to explore the generalization of training from nonsense to real words that were presented only during pre- and post-training tests. Results showed that, although both groups showed steady and substantial improvement for all vowels in nonsense words, generalization to real words was inconsistent across lis-

tener groups, different vowels, and talkers. These results indicated that: (1) native vowel phonology had a considerable influence on the success of our structured, intensive identification training protocol using many vowels; (2) nonsense word training incorporating high natural variability across phonetic contexts and talkers may not be sufficient to achieve the desired generalization to real words. These results are explained in part by the vowel acoustics (formant frequencies and vocalic duration) of the stimulus materials used for training and tests. [Work supported by NIDCD-006313 and 02229.]

3aSC5. An analysis of the factors reading rate, vocabulary ability, and speaking proficiency in relation to the test of English for international communication (TOEIC) scores of Japanese learners of English. Keiko Asano and Michiko Mochizuki Sudo (Juntendo Univ., 1-1 Hiragagakuendai Inbamura, Inbagun, Chiba 270-1695, Japan, keiko_asano@sakura.juntendo.ac.jp)

This study investigated the relationships between the TOEIC scores and the two factors of the reading rates and vocabulary abilities of Japanese learners of English. Also studied are the relationships between the scores of the two TOEIC sections and the learners' speaking proficiencies. Applying the Rauding theory, which reported the same comprehension processes underlying reading and listening, to second language acquisition, this study tested the hypothesis that the processing time for reading had an effect on that for listening. The Japanese subjects in this study took TOEIC and standardized tests of reading rates and vocabulary abilities, and also participated in a production experiment. The results showed high correlations between the TOEIC scores and the two factors of reading rates and vocabulary abilities. They also showed the tendency that the improvement of reading rates and vocabulary abilities contributed to the improvement of not only the scores of the TOEIC Reading Section, but also those of the TOEIC Listening Section. In addition, this study clarified the degree to which the speaking proficiency measured by naturalness judgments of native speakers of English corresponded to the proficiency indirectly measured by the TOEIC Listening Section.

3aSC6. The long-term retention of phonetic categories by Japanese immersion graduates. Tetsuo Harada (School of Education, Waseda Univ., 1-6-1 Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan)

English-speaking children in a Japanese immersion program, in which many content subjects are taught in Japanese, were reported to be successful in establishing phonetic categories for single and geminate stops in Japanese, although their categories deviated from those of Japanese monolinguals. However, it is unknown to what extent those children can retain their phonetic categories after they exit the program. Ten participants who had attended a Japanese immersion program in the United States from kindergarten or grade 1 to grade 5 and who were taking a university-level intermediate or advanced Japanese course at the time of data collection were recorded producing Japanese words including both singletons and geminates. Mean closure duration values were obtained, and the ratio of geminates to singletons for the immersion graduates was compared with that of English learners of Japanese having the same proficiency level in a regular Japanese program. The results show that the immersion graduates, who had been exposed to a substantial amount of Japanese at an early age, still retained the phonetic categories and could contrast singletons with geminates in Japanese, although their closure duration tended to be longer than the norm of monolingual Japanese speakers.

3aSC7. The effects of training context on perceptual learning of accented speech. Jessica E. Duke and Sabrina K. Sidaras (Dept. of Psych., Emory Univ., Atlanta, GA 30322)

This study investigates factors influencing adult listeners' perceptual learning of talker-specific and accent-general properties of spoken language. Exposure to foreign accented speech under different training con-

ditions was used to examine mechanisms involved in perceptual learning. Adult native speakers of American English were exposed to English words produced by six native-Spanish-speaking adults. During training, the Spanish-accented words were either grouped by speaker, by word, or were randomized with respect to speaker and word. The different types of training were hypothesized to direct attention to particular characteristics of the accented speech. At test, listeners were asked to transcribe novel words produced by familiar or unfamiliar accented speakers. Regardless of training condition, listeners who had been exposed to Spanish-accented speech during training were more accurate at test than an untrained control group. However, high variability training in which words were randomized led to better transcription performance at test than training with items grouped by speaker or by word. These findings suggest that high variability training may direct attention to the accent-general properties of speech necessary for perceptual learning of accented speech.

3aSC8. Development and validation of an automatic spoken Spanish test.
Nicole Wilson, Elizabeth Rosenfeld, and Jared Bernstein (Ordinate Corp., 800 El Camino Real, Menlo Park, CA 94025)

The development and validation of a fully automatic computer-based test of spoken Spanish is described. The test takes between 15 and 20 min to complete and comprises readings, elicited imitations, opposites, short-answer questions, sentence constructions, and story retellings. The only response modality is spoken Spanish. Sets of vetted items were assembled into tests presented to over 1000 adult non-native Spanish learners and over 1000 native Spanish speakers from several countries. Expert human judgments of the non-native responses showed that the spoken response material carried sufficient information for highly reliable judgments of proficiency. In the development and validation process, over 60 000 responses were transcribed and over 21 000 human judgments were analyzed. The automatic scoring system is described. Its validation includes comparisons of machine scores with interactive human interviews and human ratings of speech recordings. Results suggest that machine scores have virtually the same information that is found in Oral Proficiency Interviews conducted by the U.S. Government raters or by official ACTFL interviews. Correlation between machine scores and combined interview scores ($r = 0.92$) is higher than the interrater reliability of the professional interviewers themselves.

3aSC9. French conversations do not show evidence of phonetic accommodation to a non-native interlocutor.
Caroline Smith (Dept. of Linguist., Univ. of New Mexico, MSC 03 2130, 1 Univ. of New Mexico, Albuquerque, NM 87131-1196)

Although a speech style known as foreigner talk has been described, speech addressed to a reasonably fluent non-native interlocutor may show little or no accommodation by native speakers. This issue is of interest in understanding whether on-line modifications to speech are an important contributor to phonetic variation. Ten native speakers of Parisian French were recorded individually as they performed map tasks, one with another native speaker of French, and another with an American who identified himself as such at the beginning of the conversation, but who spoke French during the task. Previous work has found that French listeners who heard brief extracts from these conversations could not reliably identify which interlocutor was involved. In an attempt to identify other possible differences in speech to natives and non-natives, two prosodic characteristics of the conversations were analyzed, vowel devoicing and the epenthesis of word-final schwas. The hypothesis was that vowel devoicing, which is stylistically marked, would be disfavored in speech to non-natives, while schwa epenthesis, characteristic of a deliberate speech style, would be more frequent. The speakers actually showed similar patterns with both interlocutors, suggesting that variation in these properties is not conditioned by the linguistic status of the interlocutor.

3aSC10. Acoustic correlates of accentedness and intelligibility of Spanish-accented English vowels.
Lynne C. Nygaard, Sabrina K. Sidaras, Jessica E. Duke, and Stig T. Rasmussen (Dept. of Psych., Emory Univ., Atlanta, GA 30322)

The present study investigated the acoustic characteristics of English vowels produced by native speakers of Spanish and the relationship between the acoustic properties of non-native vowel production and perceived accentedness and intelligibility. Twelve native speakers of Spanish were recorded producing monosyllabic words in list format. Individual vowel tokens from a variety of consonant environments were identified and acoustic analyses were conducted, including measures of duration and formant (F1, F2, F3) frequency. Listener ratings of accentedness and intelligibility were obtained from a separate set of sentence-length utterances. The analyses showed that average overall vowel duration as well as the duration of individual vowels was positively correlated with judgments of accentedness and negatively correlated with overall sentence intelligibility. Speakers who were judged to be more accented and less intelligible produced vowels of longer duration than speakers who were judged to be less accented and more intelligible. Analyses of spectral characteristics and comparisons to native English speakers' vowel productions will also be reported. These findings suggest that the temporal properties are one aspect of non-native vowel productions that significantly contribute to the perception of accentedness as well as to the overall intelligibility of individual native Spanish speakers.

3aSC11. Factors influencing L2 vowel perception and production.
Karece Lopez (Dept. of Speech, Commun. Sci., & Theatre, St. Johns Univ., 8000 Utopia Pkwy, Jamaica, NY 11439), Yan Helen Yu (CUNY Grad. School, New York, NY), and Fredericka Bell-Berti (St. Johns Univ., Jamaica, NY 11439)

This study investigates relations among factors expected to influence vowel perception and production in second language learning, including AoA, LoR, L2 education (before coming to and in the US), and language use (in the US). Vowel production will be examined through phonetic transcription, listener identification, and acoustical analysis. Preliminary evidence indicates that, even though our participants arrived in the US after age 12, AoA is still of critical importance in second language learning. Furthermore, and as expected, LoR also contributes to L2 proficiency. AoA and LoR, however, do not account for the ability of all the participants to identify and produce the vowels of English; thus, we must also consider other factors, including L2 education and use. The hypothesis is that L2 educational background and language use influence L2 English speakers' perception and production of English vowels.

3aSC12. Production of English stressed vowels and interstress intervals within and between words in first and second language acquisition.
Michiko M. Sudo and Ikuyo Kaneko (School of Health and Sports Sci., Juntendo Univ., 1-1 Hiragagakuendai, Inbamura, Inbagun, Chiba, 270-1695, Japan, mmsudo@polka.plala.or.jp)

The purpose of this study was to observe the differences between first and second language acquisition in English durational patterns with respect to interstress intervals (ISIs). Production experiments were carried out to investigate the durational patterns used by three groups of speakers: adult native speakers of American English, American third graders, and Japanese learners of English. Two sets of English sentences were devised as the linguistic material, and the sentences in each set differed in the number of nominally unstressed syllables that intervened between a target stressed syllable and the next stressed syllable. These sets of sentences contained two different target stressed vowels, and the ISIs within and between words. The shortest ISIs were produced by adult American speakers regardless of the number of unstressd syllables in the target ISIs. The Japanese learners produced shorter ISIs than the American children in one set of sentences, while in the other set, the American children produced

shorter ISIs than the Japanese. Regarding the durations of target stressed vowels, foot-level shortening was observed in the three subject groups, the American children exhibiting the highest percentage of shortening followed by the American adults and then the Japanese.

3aSC13. Production and perception of American English vowels by native Japanese speakers. Takeshi Nozawa (Ritsumeikan Univ., 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan), Elaina M. Frieda (Auburn Univ., Auburn, AL 36849), and Ratree Wayland (Univ. of Florida, Gainesville, FL 32611-5454)

Native speakers of Japanese (NJ) produced American English vowels in /CVC/ contexts. Twelve native speakers of American English listened to and identified the vowels in the Japanese speakers' productions. In general, the vowels difficult for the NJ subjects to disambiguate in production were also difficult for them to differentiate in perception. An acoustic analysis revealed that the NJ speakers' attempted /i/ and /ɪ/ tokens and /ɑ/ and /ʌ/ tokens largely overlapped in the vowel space, but significant differences in duration were observed between their /i/ and /ɪ/ tokens. These findings agree with the results of our previous studies. In our perception experiment, /ɑ/-/ʌ/ was the most difficult vowel contrast for the same group of NJ speakers to discriminate, and longer /i/ tokens were more correctly identified and shorter /i/ tokens were misidentified more often as /ɪ/. In contrast to the perception experiment, the intelligibility of American English vowels produced by the NJ subjects was not affected by the place of articulation of the preceding consonant or by the voicing status of the following consonant. However, the American English vowels produced by the NJ subjects were less intelligible when a liquid /l/ follows, which may be attributed to their inability to produce /l/ authentically.

3aSC14. Durational characteristics in production of English consonant sequences by Japanese speakers. Kuniyoshi Tanaka (Grad. School of Humanities, Hosei Univ., 2-15-2 Ichigayatamachi, Shinjuku-ku, Tokyo 162-0843, Japan, tanaka22@abox.so-net.ne.jp) and Keiichi Tajima (Hosei Univ., Chiyoda-ku, Tokyo 102-8160, Japan)

Japanese has more limited consonant sequences (e.g., "honda," "nissan") than English (e.g., "script"). This difference becomes a problem for Japanese learners of English when they produce English words that contain consonant sequences. Japanese speakers often avoid these sequences by inserting epenthetic vowels between consonants or following word-final consonants (e.g., English "script" as /su.ku.ri.pu.to/). The present study examines how Japanese speakers produce consonant sequences in English and how speaking rate affects their productions. A group of native Japanese speakers and a control group of native English speakers read English words at normal and slow speaking rates. Segment durations in target consonant sequences were measured, and ratios of segment to word duration and segment to consonant sequence duration were calculated. Preliminary results suggest that the segment durations and ratios are clearly different between the two groups of speakers. For example, in /s/ + consonant sequences, Japanese speakers tend to produce a relatively shorter /s/ and a longer consonant compared to native English speakers. Also, the duration ratios do not seem to vary depending on the speaking rate for either group of speakers. Results will be discussed in terms of timing control in second-language learning. [Work supported by MEXT.]

3aSC15. The speech learning model: English and Japanese voice onset time in mature bilingual talkers. Timothy J. Riney (Intl. Christian Univ., Tokyo, 181-8585, Japan)

With reference to the Speech Learning Model (Flege, 1995 version), this study looked for evidence among mature American and Japanese bilinguals that (a) L1 VOT changed over time and (b) despite minimal L1-L2 phonetic differences, L2 VOT was different from L1 VOT. Five L1 English Americans and 5 L1 Japanese read the same lists of English and Japanese words containing word initial /p/, /t/, and /k/. In all, 880 tokens

(10 talkers ×2 languages ×11 words ×4 tokens) were examined. For both groups and individuals, based on measures of difference and comparisons with age-matched monolinguals (5 American and 13 Japanese), the findings were (a) slight evidence of L1 VOT change, but (b) robust evidence of new L2 categories being formed. By far the most common bilingual pattern observed was L2 compromise VOT, in which the L1 VOT remained close to monolingual VOT and the L2 VOT was positioned about halfway between that of two age-matched monolingual comparison groups.

3aSC16. Phonetic disambiguation by Korean non-native speakers of English. Jiwon Hwang and Giorgos Tserdanelis (Dept. of Linguist., Stony Brook Univ., NY 11794, gtserdanelis@sunysb.edu)

This study examined the effect of pragmatic contrast and auditory priming on the production of contrasts such as "mob∼mop" and "pat∼pet" by 22 Korean non-native speakers of English. Korean speakers played a card arrangement game with a native speaker of English (a confederate). In the pragmatic condition, critical items were arranged to enable the comparison of the same word (e.g. "mob") in two contexts, one where the situation might force subjects to make the contrast of mob and mop and one where the situation facilitates the confederates recognition of mob because there is no contrasting form. An additional auditory priming condition was induced by the confederate saying a phonetically similar word such as "what is below hob?" before the target "mob." Koreans produced more English-like forms when they had to be distinguished from the contrastive item or when they had been auditorily primed. For both b/p and ae/E contrasts, vowel duration was used to signal the distinction in both conditions. Additionally, for the b/p contrast, closure voicing percentage was significantly longer but only in the pragmatic condition. Overall both conditions have similar disambiguating effect, but have different specific consequences for cueing consonant versus vowel contrasts.

3aSC17. Phonetic similarities between English and Korean sibilants: Implications for second language acquisition. Sang Yee Cheon (Dept. of East Asian Lang. and Lit., Univ. of Hawaii at Manoa, 1890 East West Rd., #352, Honolulu, HI 96822)

Recent models of L2 acquisition suggest that foreign accent is determined by the degree of articulatory, acoustic, and perceptual similarity between L1 and L2 sounds. This study examined cross-language perceptual and acoustic similarities between Korean and English voiceless sibilant fricatives: Korean lax aspirated /s/ and tense unaspirated /s*/ versus English alveolar /s/ and palato-alveolar /ʃ/. AX discrimination tests revealed that listeners perceived the non-native contrast in terms of their L1 sound system. To determine acoustic similarity among the fricatives, two acoustic parameters were measured: duration and spectral peak frequency. It was expected that the more acoustically similar two sounds were, the more they would be perceived as the same sound by listeners. However, disparate results were obtained for perceived similarity and acoustic similarity. For instance, Korean lax /s/ and English /ʃ/ before /i/ are acoustically similar, but listeners were easily able to discriminate between them. Perceptual similarity predicts ease of acquisition better than acoustic similarity because a listener's perceptual system incorporates how the sounds are organized in their L1 phonological system. Language teachers need to consider not just degree of similarity between pairs of sounds, but also the differences between the L1 and L2 sound systems as a whole.

3aSC18. Cross-language speech sounds: Differential sensitivity and L2 accent. Miles Munro (Dept. of Cognit. Sci., Univ. of California, Irvine, 3151 Social Sci. Plaza A, Irvine, CA 92697-5100, mmunro@uci.edu)

Individuals vary in their ability to perceive phonetic differences between speech sounds. Individuals also vary in their aptitude for acquiring accurate pronunciation of a second language. The research to be presented

investigates a probable link between the two with the aim that capacity for phonetic perception may constitute a measure of second language aptitude. In a series of tasks manipulating memory and attention load, Native Mandarin speakers for whom English is a second language were tested on their ability to discriminate non-native speech contrasts that cluster around the Mandarin and English phonetic spaces with varying "distances." Each contrast pair is phonemically meaningful in a significant portion of the world's languages (UPSID 451). These scores were then related to measures of degree of accent in English. Related results examining working memory span will also be reported. Accordingly, an explanation of variability in later-learner L2 accent strength is explained as a consequence of idiosyncratic ability to perceive phonetic differences between speech sounds. This cognitive correlate's account is consistent with the notion that a greater ability to perceive small differences between novel phones is crucial to their accurate acquisition and reproduction.

3aSC19. Performance of auditory search with Japanese speech by native speakers and its learners as L2. Noriaki Asemi (Grad. School of Sci. Technol., Kobe Univ., Rokkodai 1-1, Nada-ku, Kobe, 657-8501 Japan, asemi@kobe-u.ac.jp), Ryoko Hayashi (Kobe Univ., Kobe, Japan), and Kazuhiko Kakehi (Chukyo Univ., Toyota, Japan)

Auditory search experiments using normal Japanese speech (NJS) and time-reversed speech (TRS) [Asemi *et al.*, Acoust. Sci. Technol., **24**, 145–147 (2003)] demonstrated fast and accurate judgment of the presence or absence of the target NJS among distracting TRSs. For application of an auditory search to measurement of the ability of auditory comprehension of foreign language, the effects of the learning proficiency of Japanese on the speed and accuracy in the detection of the target NJS were examined through auditory search experiments for native and non-native speakers of Japanese. The experiment comprised two parts: an NJS was the target and TRSs were distractors and vice versa. Subjects were requested to judge the presence or absence of a target speech correctly and as quickly as possible when a target and several distracters were presented simultaneously and diotically. Results showed remarkable differences between the performance of natives and non-native speakers in error rates rather than in the reaction times. Moreover, significant correlations were found between the error rates and the learning experiences only when fewer than three utterances were presented simultaneously. [Work supported by MEXT.]

3aSC20. Neurophysiological basis of temporally cued phonetic contrasts in Japanese and American English listeners. Miwako Hisagi, Valerie Shafer (CUNY-Grad. Ctr., Dept. of Speech and Hearing Sci., 365 Fifth Ave., New York, NY 10016, mhisagi@hotmail.com), and Elyse Sussman (Albert Einstein College of Medicine of Yeshiva Univ., New York, NY)

The present study examined American English (AE) listeners' perception of Japanese (JP) temporally cued contrasts of vowels and consonants. The mismatch negativity (MMN) component of event-related potentials (ERPs) was used as a measure to compare passive (visual-attention condition) and active (auditory-attention condition) discrimination in an oddball paradigm. JP nonsense temporally cued vowel ([tado] vs [taado]) and consonant contrasts ([miʃi] vs [miʃʃi]) were presented to JP listeners, for whom the temporal cues are contrastive, and to AE listeners who have no knowledge of JP. The following results were obtained. (1) With respect to the type of contrast, MMNs were larger for vowel contrasts than consonant contrasts. (2) In the auditory-attention condition both groups showed a robust MMN to vowels. (3) In the visual-attention condition JP also showed a robust MMN, while AE showed a smaller amplitude or absent MMN. This group difference was significant for the vowel contrast. (4) Interestingly, in the auditory-attention condition, AE showed a larger late negativity at frontal sites for vowel contrasts than JP. The results support the hypothesis that first-language perception is an overlearned process that may interfere with second-language speech perception. [Work supported by NSF.]

3aSC21. Vowel perception in second language acquisition. Mariche Garcia Bayonas (Dept. of Romance Lang., at Greensboro, Univ. of North Carolina, 2314 HHRA Bldg., UNCG, P.O. Box 26170, Greensboro, NC 27402-6170)

This study investigates the perception of ten English and five Spanish vowels by native-speakers (NS) and learners (L) at two levels of acquisition, and compares these vowels cross-linguistically. Research on the acquisition of vowels indicates that learners can improve their perception with exposure to the second language (L2) [Bohn and Flege (1990)]. Using Flege's (1995) Speech Learning Model as a theoretical framework, this investigation incorporated a method of adjustment (MOA) task [Johnson *et al.*, (1993)] used to analyze NS perception and production data, to add the dimension of the L2 learner data. The research question was: Are vowels selected differently by NS and L in an MOA task depending on level of proficiency? Spanish learners ($n = 40$) and English learners ($n = 80$) completed MOA tasks in which they were exposed to 120 synthetically produced vowels to analyze spectral variation in the acquisition of both sound systems, and how the learners vowel system may vary from that of the NS. In the MOA tasks they were asked to select which synthesized vowel sounds most resembled the ones whose spelling was presented to them. Results indicated that vowels are selected differently by NS and L, but there are differences among vowels.

3aSC22. Multidimensional category-goodness-difference assimilation of Canadian English /i/ and /ɪ/ to Spanish /i/. Geoffrey Stewart Morrison (Dept. of Linguist., Univ. of AB, Edmonton, AB, T6G 2E7, Canada, gsm2@ualberta.ca)

Morrison [J. Acoust. Soc. Am., **117**, 2401 (2005)] theorized that L1-Spanish learners of English initially identify the Canadian English /i/-/ɪ/ contrast via a multidimensional category-goodness-difference assimilation to Spanish /i/, with more-Spanish-/i/-like vowels (shorter vowels, lower F1 and higher F2) labelled as English /ɪ/ and less-Spanish-/i/-like vowels (longer vowels, higher F1 and lower F2) labelled as English /i/. The L1-Spanish listeners' use of duration is positively correlated with L1-English speakers' productions, but their use of spectral cues is negatively correlated; hence, exposure to English results in increased use of duration cues and reduced use of spectral cues, leading to unidimensional duration-based perception. Since, in L1-English speakers' productions, spectral properties are partially correlated with duration properties, L2-English learners can use duration-based perception as a bootstrap for learning L1-English-like spectral-based perception. Results from a new study are presented. Unlike the previous study, synthetic stimuli in the new study included vowel inherent spectral change (VISC). Results were consistent with the theory above: Vowels that were poor matches for Spanish /i/ because they had converging-VISC were labelled as English /i/, and vowels with no VISC were labelled as English /ɪ/. [Work supported by SSHRC.]

3aSC23. Multiple cue interpretation in non-native phoneme perception: Korean listeners' perception of English /s/ in varied prosodic contexts. Miran Kim (Dept. of Linguist., SUNY at Stony Brook, Add., S201, SBS Bldg, Stony Brook, NY 11794-4376, mrkim@ic.sunysb.edu)

This paper investigates prosodic influences on the non-native perception of English /s/ by Korean listeners, whose phonemic system includes a two-way distinction of fricative, plain /s/ vs. fortis /s*/. English stimuli varied in lexical stress and prosodic domain appear to be relevant to the Korean listeners perceptual responses such that more /s*/ responses to stimuli are obtained with more strengthening effects such as lexical stressed and phrase boundary initial position. Analyses of the relation

between the spectral properties of the stimuli and perception shows that prosodic effects on /sV(C)/ syllables are compatible with the characteristics that have been reported as critical attributes in the distinction of plain and fortis fricatives in Korean. Spectral measurements include F0, H1-H2 (dB), and H1-F2 (dB) values of postfricative vowels, together with the spectral characteristics of frication noise. The relation between prosodic strength in the stimuli and Koreans /s-s*/ perceptual responses implies that Korean listeners interpret prosodic information associated with English /s/ as a part of the segmental properties of English /s/ instead of attributing them to prosodic effects. This leads us to conclude that phonetic information is not limited to a segment-to-segment or prosody-to-prosody transmission, but it can be overlapped in multiple phonetic dimensions.

3aSC24. Identification of English consonants by Korean students.
Byunggon Yang (English Education Dept., Pusan Natl. Univ., 30 Changjundong, Keumjunggu, Pusan 609-735, South Korea)

Previous studies have shown that Korean students have difficulty identifying some English consonants that are not in the Korean sound inventory. The aim of this study was to examine the accuracy score of English consonants correctly identified by 130 college students in order to find out English consonants with listening problems for the Korean students. The subject's task was to identify one of the minimal pairs played in a quiet laboratory classroom. One hundred minimal pairs consisted of syllables with various onsets or codas: stops, fricatives, affricates, liquids, and nasals. Results were as follows: First, types of the correctly identified English consonants varied according to the level of each student's total score. Second, the lower the group level, the lower the accuracy score for each consonant group divided by its manner. Third, those consonants in coda position of the test words were more difficult for the students to identify than those in onset position. Fourth, the subjects showed lower accuracy in the perception of any contrast with a fricative onset or coda. Finally, the students still had problems telling voiced from voiceless consonants, especially in coda position. Future studies will focus on examining students' production data to trace their listening problems.

3aSC25. Identification of English /r/ and /l/ in noise: The effects of baseline performance. Kazuo Ueda (Dept. Appl. Inf. and Commun. Sci., Kyushu Univ., 4-9-1 Shiobaru, Minamiku, Fukuoka 815-8540, Japan, ueda@design.kyushu-u.ac.jp), Ryo Komaki, and Takahiro Adachi (ATR, Soraku-gun, Kyoto 619-0288, Japan)

The effects of baseline performance on identification of English /r/ and /l/ in noise by Japanese listeners were examined. Japanese and American listeners' perception of /r/ and /l/ was measured under various signal-to-noise ratios. Each Japanese listener's baseline performance was also observed with an identification test with a large set of stimuli. Generally, the signal-to-noise ratio had a similar effect across Japanese listeners of different baseline performances, although the overall accuracy was dominated by the baseline. When 15 days of /r/-/l/ identification training in quiet were applied to a group of Japanese listeners, the training effect generalized to identification performance in noise. However, the performance of Japanese listeners did not reach the level that the native listeners exhibited. [Work supported by JSPS and 21st Century COE, Kyushu University.]

3aSC26. Effects of learners' L2 proficiency and acoustic/semantic contexts of stimuli on the phoneme identification training. Yuko Ikuma and Reiko Akahane-Yamada (ATR, 2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan)

Former studies investigated that word recognition by L2 learners is affected by both acoustic and semantic contexts [Rothwell and Akahane-Yamada, J. Acout. Soc. Am. **112**, 2386 (2002)]. It was also shown that these two context effects lead to different achievement in perception train-

ing [Ikuma and Akahane-Yamada, J. Acout. Soc. Am. **115**, 2393 (2004)]. In this study, native speakers of Japanese were trained to identify English /r/ and /l/ within words or sentences. L2 proficiency level for each trainee was also measured by a short written test regarding vocabulary and grammar. Results showed that training with semantically contextual stimuli improved the identification accuracy for stimuli with semantically contextual cues, but not for stimuli without semantic cues. Furthermore, the L2 proficiency level correlated only with the improvement in identification accuracy of target phonemes embedded in the semantically contextual sentences. These results suggest that L2 learners tend to use semantic cues rather than difficult acoustic cues. However, semantic context may moderate the effect of auditory perception training. Thus, training with semantically neutral stimulus is necessary to improve perception ability even for high proficiency level learners. [Work supported by JSPS.]

3aSC27. Asymmetry in discrimination of an American English contrast, /l-r/, by adult Japanese learners of English. Teruaki Tsushima (Ryutsu Kagaku Univ., 3-1, Gakuen-Nishi-Machi, Nishi-Ku, Kobe, 651-2188, Japan)

The present study provided further evidence on an asymmetry in discrimination of an American English contrast, /l-r/, by Japanese listeners. Experiment 1 ($N=14$) examined stimulus order effects in discrimination of the contrast in the AX discrimination procedure with an ISI of 300 and 1500 ms, using naturally produced stimuli. The results showed discrimination accuracy was significantly lower when /l/ was followed by /r/ than otherwise, under the ISI of 1500 ms. Experiment 2 ($N=17$) examined whether discrimination significantly differed when a typical /l/ was paired with an ambiguous /r/ and vice versa, using synthesized stimuli from a ten-step continuum (i.e., stimulus 1-7 and 4-10). The participants were tested on their identification of the stimuli, and discrimination of the stimulus pairs with the ISI of 300 and 1500 ms. The discrimination sensitivity (i.e., d'), based on the identification performance, was not significantly different between the two pairs. In the discrimination test, however, the sensitivity was significantly lower in the pair including the typical /l/ than in the other, under the ISI of 1500 ms. The results support the hypothesis that categorization of a relatively nativelike speech stimulus (i.e., /l/) in a discrimination task may induce some perceptual assimilation process.

3aSC28. Context really matters!—The influence of input types on L2 speakers' perception of English liquids. Sally Chen and Janice Fon (Grad. Inst. of Linguist., Natl. Taiwan Univ., 1 Roosevelt Rd., Sec. 4, Taipei 106, Taiwan)

This study aims to investigate how different input types can help improve the perception of language learners on L2 sounds. Twenty-four college students of a remedial English class are divided into three groups and are given audio inputs of different types. The first group listens to a wordlist, the second group to a story incorporating all the words on the wordlist, and the third group to a piece of classical music, serving as the control. The wordlist consists of 56 monosyllabic English words with word-initial liquids or liquid-containing clusters. Before the listening task, all three groups are given an online l/r perception test containing all the words on the wordlist (pretest). After the listening task, the same perception test is again administered, including the original 56 words and another 56 phonotactically matched pseudo-V words (posttest). Both accuracy and reaction time are measured for the two perception tests. Preliminary results showed that all participants had significantly higher accuracy and shorter RT for real words in the posttest. The story group showed the biggest improvement for both measures, followed by the wordlist group. The control showed the least improvement.

3aSC29. Categorization of a sibilant+vowel continuum in Japanese, Mandarin, and English. Terrance M. Nearey, Aya Okamoto, Chunling Zhang, and Ron I. Thomson (Dept. of Linguist., Univ. of AB, Edmonton, AB, T6G 2E7, Canada)

A continuum of 168 synthetic sibilant+vowel stimuli was categorized by listeners of three languages (Japanese, Mandarin, and English). The contoid portion of each stimulus consisted of a 190-ms high-pass filtered white noise whose cutoff frequency varied from 1800 to 3800 Hz in 7 steps. (Signals were low-pass filtered at 8500 Hz before resampling to 22.05 kHz for playback.) Each (cascade formant) vocoid portion was 210 ms in duration varying in three $F1$ levels (from 300 to 430 Hz) crossed with four $F2$ levels (from 850 to 2100 Hz, with fixed $F3$ through $F5$). Each of the 84 noise+vocoid patterns was combined with two different $F2$ transition patterns, designed to slightly favor more alveolar or more palatal fricative responses. Appropriate response sets were determined in pilot studies for each language by native speakers on the research team. Corresponding software was designed to allow computer-controlled categorization. Data collection for 9 Japanese, 20 Mandarin, and 4 English listeners is complete (more is planned). Responses to the stimuli spanned at two or three sibilant and several (high-to-mid) vowel categories in each language. Possible effects of language-specific phonotactics on the response patterns will be discussed.

3aSC30. Perceptual cue learning. Grant McGuire (The Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, mcguire@ling.osu.edu)

This paper describes results from an on-going perceptual training study in which native listeners of English are trained perceive a categorical difference between Polish alveopalatal and retroflex sibilants. The stimuli consist of naturally produced, modified two-dimensional stimuli varying by fricative noise and formant transition cues. Subjects are trained to categorize using only the fricative dimension, formant transition dimension, or both dimensions. Testing consists of a two-alternative forced-choice discrimination paradigm and stimulus labeling. Initial results from pilot work and subjects in the fricative dimension training group demonstrate that some listeners can easily attend to either dimension independently for categorization where training focuses attention and sharpens boundaries while others rely heavily on the vocalic cues, only attending reliably to frication noise cues with several training sessions or through verbal instruction. A handful of subjects did not show any improvement due to training; most of these subjects also showed no initial categorization strategy. Moreover, sensitivity to the fricative dimension and boundary increases for all learners (measured in d'), with no corresponding increase or decrease in sensitivity to vocalic cues. Further results from other training conditions will also be reported. [Work supported by NIH-NIDCD.]

3aSC31. Cantonese and Japanese listeners' processing of Russian onset and coda stops. Alexei Kochetov (Dept. of Linguist., Simon Fraser Univ., 8888 Univ. Dr., Burnaby V5A 1S6, Canada, alexei_kochetov@sfu.ca), Connie K. So (Univ. of Western Sydney, Australia), and Nicole Carter (Simon Fraser Univ., Burnaby V5A 1S6, Canada)

The study investigated influences of native phonological and phonetic knowledge on Cantonese and Japanese listeners' processing of non-native stop consonants. Phonologically, Cantonese Chinese has a three-way place contrast in stops in onset and coda positions; Japanese has a similar contrast, however, only in onset position. Phonetically, Cantonese coda stops are acoustically unreleased; Japanese stops followed by devoiced vowels are acoustically similar to released coda stops in other languages, such as Russian. Two groups of listeners, native speakers of Hong-Kong Cantonese and Japanese, were presented with sequences of Russian voiceless stops (VC1#C2V, where C1/C2=/p/, /t/, or /k/). Two tasks were employed: (*i*) identification of onset or coda stops and (*ii*) discrimination of sequences that differed in either onset or coda consonant. Both groups

performed equally well in the identification and discrimination of onset stops. However, Japanese listeners performed significantly better than Cantonese listeners in the identification and discrimination of coda stops. The findings suggest that in non-native listening, the low-level native phonetic knowledge—the acoustic realization of stop place contrasts—can override the higher-level phonological knowledge—syllable structure constraints on the distribution of place features. [Work supported by Social Sciences and Humanities Research Council of Canada.]

3aSC32. Relationships between the intelligibility and acoustic properties of native English and Mandarin-accented English. Rachel Hayes-Harb (Dept. of Linguist., Univ. of Utah, 255 S. Central Campus Dr., Rm. 2330, Salt Lake City, UT 84112), Bruce Smith (Univ. of Utah, Salt Lake City, UT 84112), Tessa Bent (Indiana Univ., Bloomington, IN 47405), and Ann Bradlow (Northwestern Univ., Evanston, IL 60208)

Studies examining effects of talkers' and listeners' linguistic backgrounds on speech intelligibility have produced mixed results, and relatively little is known about which acoustic-phonetic features of native and non-native speech are responsible for many findings. These issues are addressed here via investigation of how acoustic properties of final consonant voicing contrasts influence word intelligibility. Native English and native Mandarin talkers' productions of English words were presented in isolation to native English and native Mandarin listeners in an identification task. Intelligibility findings were related to preceding vowel duration, stop consonant duration, burst duration, and proportion of voicing during stop closure. As expected, native English listeners were more accurate identifying words spoken by native than non-native talkers. In contrast to previous findings of an interlanguage speech intelligibility benefit (ISIB) for sentences, non-native listeners found native speech more intelligible than non-native speech. However, non-native listeners found non-native speech more intelligible than did native listeners. For native talkers, all four acoustic measures contributed significantly to correct identification of voicing contrasts by both native and non-native listeners (with closure voicing most salient). In contrast, for non-native talkers, listeners tended to rely primarily on closure voicing and burst duration. Ramifications for ISIB will be discussed.

3aSC33. English listeners perceptual assimilations for Zulu sounds: Evidence in support of the articulatory organ hypothesis. Andrea Levitt (Wellesley College, 106 Central St., Wellesley, MA 02481, and Haskins Labs., 300 George St., New Haven, CT 06511, alevitt@wellesley.edu), Catherine Best (Univ. Western Sydney, Penrith South NSW 1797, Australia), Louis Goldstein (Yale Univ., New Haven, CT), and Angela Carpenter (Wellesley College, Wellesley, MA 02481)

The articulatory organ hypothesis predicts that listeners' favored perceptual assimilations for non-native sounds will be native sounds that are produced by the same articulatory organ [L. Goldstein and C. Fowler in *Phonetics and Phonology in Language Comprehension and Production*, 159–207 (2003)]. Using discrimination paradigms, we tested English-speaking, adult listeners' perceptual assimilations for four non-native Zulu consonants to see whether assimilations are stronger for sounds articulated with the same vocal-tract organ or those with the same constriction degree. For each Zulu sound, three possible English-language assimilations were provided: a best match, based on the most frequent consonant assimilation provided by the subjects in a previous study [C.T. Best, G.W. McRoberts, and E. Goodell, J. Acoust. Soc. Am. **109**, 775–794 (2001)], an assimilation match where the consonant uses the same articulatory organ, and one where the consonant uses the same constriction degree as the Zulu target. For example, for the bilabial implosive, the best match was /b/ (same articulator and constriction degree), the same organ match /w/, and the same constriction match /d/. Although there were differences in the patterns of responses to the four sounds, overall the results provided support for the articulatory organ hypothesis predictions.

3aSC34. Learning of non-native tonal contrasts with or without tonal context. Valter Ciocca (Div. of Speech & Hearing Sci., Univ. of Hong Kong, 34 Hospital Rd., Hong Kong, vciocca@hkusua.hku.hk), Alexander L. Francis, and Yanhong Zhang (Purdue Univ., West Lafayette, IN 47907-2038)

The learning of non-native (Cantonese) tonal categories by native speakers of another tonal language (Mandarin) was investigated under two learning conditions. For listeners in the tonal context (TC) group, learning occurred by listening to target tones within various carrier phrases. Listeners in the no-context group (NC) were trained using target tones produced as single words. Training stimuli were produced by the same speakers for both groups. A high-variability procedure with feedback was used for training. Sixteen listeners (eight in each group) took part in eight sessions over a period of about 1 month. Categorical perception of Cantonese tones was measured before training (following a brief familiarization session) and after training. For both groups, tone perception was tested for one level- and one contour-tone continua using identification and discrimination procedures. Identification performance before and after training did not differ between the two groups for the level-tone continuum, showing evidence of learning of the (non-native) mid-level category. For the contour-tone continuum, the TC group showed better learning of the (non-native) low-rising/low-level contrast, as well as better perceptual distinction between low-rising and high-rising tones. [Work supported by Hong Kong RGC Grant HKU 7224/03H.]

3aSC35. Native and non-native identification of acoustically modified Mandarin tones. Chao-Yang Lee (School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701), Liang Tao, and Zinny S. Bond (Ohio Univ., Athens, OH 45701)

Lexical tones have been known to be one of the most difficult aspects of tone languages to master for non-native learners. Extending upon Gottfried and Suiter's study [J. Phonetics **25**, 207–231 (1997)] of tone identification in silent center syllables, four experiments examined how well Mandarin tones can be identified with partial acoustic information and whether native and non-native listeners show differences in the use of the limited acoustic information. Twelve minimal tone pairs including all six tonal contrasts in Mandarin were digitally processed to generate four types of stimuli: silent-center, center-only, onset-only, and intact syllables. The stimuli were presented in the original carrier phrase, excised from the carrier phrase, or excised and pasted onto another carrier phrase. Participants included 40 non-native speakers with 1–3 years of Mandarin instruction at Ohio University and 40 native speakers. Both learners and native speakers could identify the tones under all modification conditions at better than chance levels. For both groups, the onset-only syllables were the most difficult. Tone 2 received the fewest correct identifications and longest response times, being confused with tone 3.

3aSC36. Acoustic correlates of English lexical stress produced by native speakers of Mandarin Chinese. Yanhong Zhang and Alexander L. Francis (Program in Linguist., Purdue Univ., West Lafayette, IN 47907)

Native speakers of Mandarin Chinese have difficulty producing native-like English stress contrasts. Acoustically, English lexical stress is multidimensional, involving manipulation of fundamental frequency (f0), duration, amplitude, and vowel quality. Errors in any or all of these correlates could result in poor realization of the stress contrast, but it is unknown which cues are most problematic for Mandarin speakers. Here we compared the use of these correlates in the production of lexical stress contrasts by 20 Mandarin and English speakers. Results showed that Mandarin speakers produced significantly less acceptable stress contrasts, although they did use all four acoustic correlates to distinguish stressed from unstressed syllables. Mandarin and English speakers' use of amplitude and duration were comparable for both stressed and unstressed syllables, but Mandarin speakers signaled stressed syllables with a higher f0 than English speakers. There were also significant differences in formant

patterns across groups, suggesting that Mandarin speakers were able to achieve appropriate vowel reduction in certain unstressed syllables, but not in others. Results suggest that Mandarin listeners' production of lexical stress contrasts in English is influenced in part by native-language experience with Mandarin lexical tones, and in part by similarities and differences between Mandarin and English vowel inventories.

3aSC37. On the characteristics of English speech timing control by Japanese learners. Hajime Tsubaki, Shizuka Nakamura, and Yoshinori Sagisaka (GITI Waseda Univ., 29-7 Bldg. 1-3-10, Nishi-Waseda, Shinjuku-ku, Tokyo 169-0051, Japan, hjm-tsubaki@asagi.waseda.jp)

This paper introduces analysis results of temporal characteristics of English language uttered by Japanese subjects aiming at automatic evaluation of L2 prosody control. As a preliminary study, 11 sentences uttered by about 200 Japanese children were compared with native's speech. The correlation between subjective scores on temporal naturalness and differences in segmental durations were measured in different speech units at every position for sentence by sentence. A wide range of correlation differences were observed such as from -0.70 to 0.23 at unstressed syllables, from -0.33 to 0.19 at stressed syllables, and from -0.70 to -0.04 at function words. The strongest negative correlation -0.70 at unstressed syllables and at function words reflects L1 characteristics of mora-timing characteristics of Japanese. Further analysis results will be given at the conference with additional huge L2 speech database currently under construction. [Work supported in part by Waseda Univ. RISE research project of "Analysis and modeling of human mechanism in speech and language processing" and Grant-in-Aid for and Scientific Research A-2, No. 16200016 of JSPS.]

3aSC38. Acquisition of Japanese pitch-accent by non-native speakers. Tomoko Kozasa (Dept. of Linguist., California State Univ., Fresno, 5245 N Backer Ave., M/S PB92, Fresno, CA 93740-8001)

Vowel duration and $F0$ movement in Japanese pitch-accented words produced by non-native speakers were investigated. Phonemic vowel length and pitch accent are two prosodic properties of Japanese that are problematic for language learners. The vowel length distinction is typically realized by durational differences, and pitch accent involves $F0$ movement from high to low within an accented word. Equal numbers of target words and filler words were produced by 50 speakers by repeating after audio recorded Japanese sentences one at a time. The stimulus sentences were uttered by a native speaker of Tokyo Japanese. Speakers were randomly divided into two groups. For group A, a visual stimulus indicating the native speaker's pitch contour for the target word appeared right after each audio stimulus. Speakers were instructed to follow the visual stimulus as they produced the target word. No visual stimulus was provided to group B. A sufficient durational distinction between long and short vowels was produced by both groups. However, group B failed to produce adequate amount or direction of $F0$ movement. These results suggest that explicitly showing pitch contour to language learners improves their pronunciation of pitch accent.

3aSC39. Native speakers and Korean-speaking advanced learners of English show similar discrimination of English prosodic contours. Hyekyung Hwang and Amy J. Schafer (Dept. of Linguist., Univ. of Hawaii, 569 Moore Hall, 1890 East-West Rd., Honolulu, HI 96822, hyekyung@hawaii.edu)

Previous studies of native versus non-native perceivers argue that phonology can shape perception (e.g., Huang, 2001; Hume et al., 1999). The phonological inventory of prosodic phrasing contours differs between Korean and English. Phonologically, Korean accentual phrases have final rises (within intonation phrases), although this varies at the phonetic level, while English intermediate phrases (ips) can end in falling, rising, or level contours. Previous work has suggested that advanced learners (L2ers) use

ips markedly less effectively than native speakers (L1ers) to recover English syntactic structure (Hwang and Schafer, 2005). This study investigated whether that is due to perceptual differences. L1ers and L2ers listened to pairs of phrases in an AX task that crossed boundary strength (same versus different size) with contour: rising (L*L*H-) versus level (L*L*L- or L*L*) or falling (H*H*L-) versus level (H*H*H- or H*H*). Little variation was found between L1ers' and L2ers' discrimination, and L2ers' discrimination patterns correlated highly with L1ers' ($r = 0.966$). For both groups a falling-level contrast was more salient than a rising-level contrast (in the context tested) and boundary strength contrasts were more difficult than contour contrasts. The results suggest that the L2ers' poor use of ips in comprehension is likely due to difficulty with prosody-syntax mappings.

3aSC40. The effect of phonological knowledge on word learning in second language acquisition. Junko Maekawa (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, 1000 Sunnyside Ave., 3001 Dole Ctr., Lawrence, KS 66045-7555, junko@ku.edu)

Past research in first language (L1) acquisition suggests that phonological knowledge (individual sounds and sound sequences) influences word learning. In second language (L2) word learning, phonological knowledge is dependent on the mapping of L2 sounds to L1. This study explores how the L2-L1 phonological mapping affects L2 word learning for native speakers of Japanese learning English. The learners phonological knowledge was explored through a perceptual discrimination task (same-different task) and a production probe containing English real words and nonwords. L2 word learning was explored through exposure to novel words and comprehension and production testing. The results of the same-different task showed that r-l and θ-s contrasts were the most difficult to perceive, supporting findings from past research. For the production probe, w, t, and s were produced more accurately than r, l, and θ in both real words and nonwords. For L2 word learning, nonwords with w, t, and s were learned more rapidly than nonwords with r, l, and θ in production, but not comprehension. Taken together, phonological knowledge appears to constrain L2 word learning for language expression, but not comprehension. This mirrors the findings for L1 acquisition.

3aSC41. Context effects in recognition of English disyllabic word and nonwords by native and non-native listeners. Robert Felty (Dept. of Linguist. and German, Univ. of Michigan, 3110 Modern Lang. Bldg., 812 E. Washington St., Ann Arbor, MI 48109-1275) and José Benkí (Univ. of Michigan, Ann Arbor, MI 48106)

30 native (L1 = English) and 30 non-native (L1 = German) speakers listened to English CVCCVC words and nonwords mixed with noise. Of the words, half were monomorphemic and half bimorphemic. The j-factor model [Boothroyd and Nittrouer, J. Acoust. Soc. Am. **84**, 101–114 (1988)] was used as a measure of lexical context effects. For both native and non-native speakers, words showed greater context effects than nonwords, though the difference was not as large for non-native speakers. Monomorphemic words also exhibited greater context effects than bimorphemic words, though only for native speakers. In an items analysis, a negative correlation with j was found for lexical frequency and a positive correlation for neighborhood density, for both native and non-native speakers, though more pronounced in native speakers. Misperceptions between native and non-native speakers were also compared, with non-native speakers showing patterns predictable by phonological differences between the two languages. The j-factor results extend previous research using CVC stimuli, providing additional support for its efficacy as a measure of context effects. The differences (and lack there of for non-native speakers) between mono- and bimorphemic words are interpreted as evidence in support of a combinatorial model of lexical access. [Work supported by NIH/NIDCD.]

3aSC42. Recognizing English words by Japanese speakers. Kiyoko Yoneyama (Dept. of English, Daito Bunka Univ., 1-9-1 Takashimadaira, Itabashi, Tokyo, 175-8571, Japan, yoneyama@ic.daito.ac.jp)

Japanese listeners are influence by word frequency and neighborhood density (Yoneyama, 2002) as in other languages such as English. Interestingly, non-native listeners are also affected by word frequency and neighborhood density when they recognize English (e.g., Bradlow and Pisoni, 1999; Imai *et al.*, 2005). Based on the previous studies of neighborhood density by non-native listeners, Yoneyama (2006) investigated whether Japanese listeners learning English employ two types of lexical information (word frequency and neighborhood density) when they recognize English words, as in Imai *et al.* (2005). The results were different from the ones of Imai *et al.* (2005). Japanese listeners were sensitive to both probabilistic phonotactics and word frequency. A strong correlation between probabilistic phonotactics and neighborhood density still predict Japanese listeners can be influenced by neighborhood density in English word recognition. A critical difference between Yoneyama (2006) and Imai *et al.* (2005) is that they did not use the exact same materials, although Yoneyama (2005) selected the words as described in a conference paper of an earlier version of Imai *et al.* (2005). This study reports the results of the experiment in which the materials of Imai *et al.* (2005) were used.

Session 3aSP

Signal Processing in Acoustics: Sensor Array and Its Applications

Henry Cox, Cochair
4350 N. Fairfax Dr., Ste. 470, Arlington, VA 22203

Masato Miyoshi, Cochair
NTT Communication Science Labs., 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

Chair's Introduction—8:00

Invited Papers

8:05

3aSP1. Application of a microphone array to the automatic structuring of meeting recordings. Futoshi Asano, Jun Ogata, and Yosuke Matsusaka (AIST, Central 2 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan)

A microphone array system applied to the automatic structuring of meeting recordings is introduced. First, speakers in a meeting are identified at every time block (0.25 s) by sound localization using a microphone array, and then the speech events are extracted (structuring). Next, overlaps and insertions of speech events such as "Uh-huh," which greatly reduce the performance of the automatic speech recognition, are separated out. The difficulty in eliminating overlaps and insertions is that the overlapping sections are sometimes very short, and data sufficiently long for conventional separation techniques, such as independent component analysis, cannot be obtained. In this system, a maximum-likelihood beamformer is modified so that all the information necessary for constructing the separation filter, such as speaker location and spatial characteristics of noise, are estimated from the recorded data based he structuring information. Evaluation of the performance of the speech event separation by automatic speech recognition is also presented.

8:25

3aSP2. Eigenbeam beamforming for microphone arrays. Jens Meyer and Gary W. Elko (mh acoustics LLC, 25-A Summit Ave., Summit, NJ 07901)

In this talk, we describe a beamforming spherical microphone array consisting of many acoustic pressure sensors mounted appropriately on the surface of a rigid sphere. The associated beamformer is implemented in a two-stage process: a first stage that decomposes of the sound-field into spatially orthonormal spherical harmonics, which are called eigenbeams, and a second stage that combines the eigenbeams to form the overall output beampattern. It is shown that this approach allows for a computationally efficient, flexible, and scalable beamformer structure. As an example of beamforming flexibility, an acoustic zoom feature will be described. This application enables continuous variability of the beamwidth in a simple and elegant manner. Simultaneously, the array beam-pattern can be steered without affecting the beampattern shape. There are many applications for such a feature, one being the coordination of the acoustic and image scenes of microphone and a video camera.

8:45

3aSP3. Effect of low independency on independent component analysis and countermeasures against it. Hidefumi Kobatake (Tokyo Univ. of Agriculture and Technol., 3-8-1, Harumi-cho, Fuchu 183-8538, Japan, kobatake@cc.tuat.ac.jp) and Xuebin Hu (Fuji Photo Film Co., Ltd., Chuo-ku, Tokyo 104-0061, Japan)

Independent component analysis (ICA) is extensively adopted in blind signal separation (BSS). It works well in general. However, processing must be made at every short-time interval and, as a consequence, the assumption of independency cannot always hold due to insufficient data length. This local low-independency can be the cause of degradation in the overall separation performance. In this paper, two approaches to solve that problem caused by local low-independency between source signals are presented. The first one is the combination of ICA-based BSS with beamformer (BF). The BF does not depend on the independence assumption. In the proposed method, ICA and BF are applied in parallel. The frequency components separated by those methods are evaluated by each other and the frequency bin with higher segmental similarity measure is adopted. The second one is a kind of TDD method with a recursive algorithm. A criterion on the performance of separation is introduced. The proposed method has a function to improve the separation filter characteristics recursively. Experiments to evaluate the performance of those methods and demonstrations are given.

9:05

3aSP4. Some current challenges in multichannel acoustic signal processing. Walter Kellermann (Univ. Erlangen Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany, wk@LNT.de)

Hands-free human/machine communication involves major challenges for today's state of the art signal processing if the humans should not be forced to carry or wear any technical gear and should be free to move. The spatial domain sampled by multiple loudspeakers for reproduction and multiple microphones for signal acquisition allows for spatialized listening experiences and for

3a THU. AM

spatially selective acoustic scene analysis, respectively. Along with the possibility to serve several human communication partners simultaneously, such systems also introduce the need for localization algorithms. Three main types of impairment have to be addressed by signal processing algorithms to fulfill the promises of the general concept: first, acoustic feedback from the loudspeakers into the microphones; second, ambient noise and undesired pointlike interferers; and, finally, echoes and reverberation as picked up by the human ear when listening to the loudspeaker signals or as recorded by the microphones that actually should record only the direct-path component of the source signal. Accordingly, some of the current challenges in multichannel acoustic echo cancellation, in spatially selective noise and interference suppression, in source separation and localization, and in dereverberation are investigated.

9:25

3aSP5. Sound source tracking for security camera using sensor array utilizing object-related transfer function. Tsuyoshi Usagawa, Toshimichi Takada, and Yoshifumi Chisaki (Grad. School of Sci. and Technol., Kumamoto Univesity, 2-39-1 Kurokami, Kumamoto 860-8555, Japan, tuie@cs.kumamoto-u.ac.jp)

A frequency-domain binaural model has been proposed as a computer model of human auditory system, and it can estimate the direction of arrival (DOA) of target sound source with only two inputs [Nakashima *et al.*, Acoust. Sci & Tech. **21**, 172–178 (2003)]. It utilizes a database of interaural level and phase differences obtained from head-related transfer function (HRTF) for DOA estimation. Since the interaural level and phase differences arise from the diffraction of head, similar ones can be obtained from the diffraction of an object, such as a transparent cover of security camera. In this paper, a source tracking method which utilizes an object-related transfer function is proposed. Three sensors attached on a dome-shaped camera cover are used, and the interchannel level and phase differences obtained from object-related transfer function are used for DOA estimation. The performance of sound source tracking is discussed based on the accuracy of DOA estimation as well as observation latency by the computer simulation. [Research partially supported by the MEXT, Grant-in-Aid for Scientific Research (C), 2006, 18500135 and carried out by the Cooperative Research Project Program of the RIEC, Tohoku University (Grant #H16.A02).]

9:45

3aSP6. Low-complexity design schemes for broadband electroacoustic arrays. Michael Goodwin (Creative Adv. Technol. Ctr., 1500 Green Hills Rd., Ste. 205, Scotts Valley, CA 95066, mgoodwin@atc.creative.com)

A variety of methods has been described in the literature for realizing broadband linear electroacoustic arrays, i.e., arrays which exhibit a nearly frequency-invariant response. For example, filter-network beamformers can be designed to achieve a specified beam pattern over a wide frequency range. Some applications, such as low-cost consumer audio products, do not involve strict target pattern requirements and call for lower complexity broadbanding solutions. In this presentation, three broadband array design approaches with low processing overhead and implementation cost are described. The key idea is to configure the array such that its response is essentially a magnification of the response of an individual element; the array geometry and beamformer are designed so as to not introduce directionality. In the first method, the array weights are an approximation of the impulse response of an all-pass filter; in the second method, the array elements are nonuniformly spaced; and, in the third method, arbitrary delays are applied to the elemental signals. Relationships to approaches previously described in the literature are discussed, optimization schemes based on minimizing the array response variation are presented, and the effectiveness of the array designs for realizing a robust broadband wide-angle response is demonstrated.

10:05–10:20 Break

Contributed Papers

10:20

3aSP7. Three-dimensional microphone array intensity vector sensors for measurement of various acoustical properties in different spatial environment. Khalid H. Miah and Elmer L. Hixson (Univ. of Texas, Austin, TX 78712, miah@mail.utexas.edu)

Precise measurement of different acoustical quantities in different spatial environment has been the key in improving sound-quality and reducing acoustical noise from various sources. In this report, a seven-element 3-D microphone array configuration has been used to measure different acoustical quantities (pressure, particle velocity vectors, intensity vectors, energy density, kinetic and potential energy) in various enclosed environments. Microphones in the array were arranged in an inner and an outer triangular fashion with a reference microphone at the center to address high frequency (1–6 kHz) and low frequency (100–1000 Hz) components of sound signals. Statistical distribution of acoustic field has been used for measurements in large and irregular-shaped room. Statistical distribution of acoustical field using a four-microphone configuration implemented in a previous study (M. W. Budhiantho, Acoustic Velocity Related Distributions dissertation, The University of Texas at Austin, August 1997) has been extended to a seven-microphone configuration in this report. An effort has been made to model acoustical field near the boundaries of an enclosed space (reverberation room). Also 3-D imaging technique of acoustic intensity is also presented in this report.

10:35

3aSP8. Acoustic source localization with eye array. Hedayat Alghassi, Shahram Tafazoli, and Peter Lawrence (Dept. of Elec. and Comput. Eng., University of British Columbia, Vancouver, BC V6T 1Z4, Canada, hedayata@ece.ubc.ca)

A novel signal-processing algorithm and array for sound source localization (SSL) in three-dimensional space is presented. This method, which has similarity to the eye in localization of light rays, consists of a novel hemispherical microphone array with 26 microphones on the shell and one microphone in the sphere center. The microphones on the shell map a geodesic hemisphere called two-frequency icosahedron; hence, each microphone has at least four other orthogonal microphones. A signal-processing scheme utilizes parallel creation of a special closeness function for each microphone direction on the shell in the time domain. Each closeness function cell (lens cell) consists of center microphone, shell microphone, and one of its orthogonals. The closeness function output values are linearly proportional to spatial angular difference between the sound source direction and each of the shell microphone directions. By choosing microphone directions corresponding to the highest closeness function values and implementing a linear weighted spatial averaging on them, the sound source direction is estimated. Contrary to traditional SSL techniques, this method is based on simple parallel mathematical calculations

in the time domain with low computational costs. The laboratory implementation of the array and algorithm shows reasonable accuracy in a reverberant room.

10:50

3aSP9. Bayesian space-time processing for acoustic array source estimation using a towed array. James V Candy (Lawrence Livermore Natt. Lab., Univ. of California, P.O. Box 808, L-156, Livermore, CA 94551) and Simon J. Godsill (Univ. of Cambridge, Cambridge CB2 1PZ, UK)

The development of a model-based Bayesian processor for a horizontally towed array using synthesized narrow-band, spatio-temporal data is discussed. The basic processor is developed from the Bayesian perspective, capturing the bearing estimation problem of acoustic sources moving in a surveillance volume. Far-field sources characterized by simplified random motion as well as nonlinear measurement models representing the array motion and propagation of the source signatures contaminated by noise are incorporated into the processor. The entire dynamical system is cast within a Bayesian framework aimed at estimating the underlying posterior filtering distribution of the bearing in a pseudo real-time environment. It is shown that the resulting sequential Bayesian processor is capable of dynamically estimating the instantaneous posterior distribution enabling the extraction of the bearing estimates. The problem solution involves the estimation of the posterior using sequential importance sampling/resampling leading to the development of a particle filter for this problem. The particle filter is simply an empirical representation of the posterior characterized by the weights (probability) and location variables (particles). It is shown how the particle filter successfully extracts the desired signals comparing the results to classical nonlinear processors (e.g., extended/unscented Kalman filters) for this problem.

11:05

3aSP10. A design of array-microphone system using pressure-gradient microphones and two-dimensional finite impulse response filters. Hiroyuki Naono (Matsushita Tech. Information Services Co., Ltd., 25-10 Nishiyama Adachi Yawata city, Kyoto 614-8351 Japan, yql07401@nifty.ne.jp) and Kiyoshi Nishikawa (Kanazawa Univ., Kakuma, Kanazawa-shi, 920-1192 Japan)

Delay-and-sum array microphones consisting of a linear array of microphones can achieve sharp directivity at high frequencies. A disadvantage is that the directivity is frequencydependent. An array microphone using a fan filter is proposed to resolve that problem. However, it is necessary to increase the filter order to form a sharp beam at a low frequency. This paper is reported improvement of the array microphone directivity, realizing a narrow beam with a low order of the filter using a two-dimensional frequency domain design. Using four microphones, a second-order pressure-gradient microphone is formed. The spacing of the microphones is set for high sensitivity and broad bandwidth in the operating frequency range. These microphone units are placed along a line and are cascaded to a two-dimensional fan filter so that the load is distributed. In

the proposed configuration, the second-order pressure-gradient microphone is cascaded with an axially symmetric fan filter. The configuration thereby forms a single broadside beam. Consequently, the directional characteristics of the pressure-gradient microphone are reflected in the array microphone and a sharp directivity, which is frequencyindependent, is obtained over a wide frequency range.

11:20

3aSP11. Pressure-sensitive paint as a distributed optical microphone array. James W. Gregory[a], John P. Sullivan (School of Aeronautics and Astronautics, Purdue Univ., 315 N. Grant St., West Lafayette, IN 47907, jim.gregory@alumni.purdue.edu), Sameh S. Wanis, and Narayanan M. Komerath (Georgia Inst. of Technol., Atlanta, GA 30332)

Pressure-sensitive paint is presented and evaluated in this paper as a quantitative technique for measurement of acoustic pressure fluctuations. This work is the culmination of advances in paint technology, which enable unsteady measurements of fluctuations over 10 kHz at pressure levels as low as 125 dB. Pressure-sensitive paint may be thought of as a nanoscale array of optical microphones with a spatial resolution limited primarily by the resolution of the imaging device. Thus, pressure-sensitive paint is a powerful tool for making high-amplitude sound pressure measurements. In this work, the paint was used to record ensemble-averaged, time-resolved, quantitative measurements of two-dimensional mode shapes in an acoustic resonance cavity. A wall-mounted speaker generated nonlinear, standing acoustic waves in a rigid enclosure measuring 216 mm wide, 169 mm high, and 102 mm deep. The paint recorded the acoustic surface pressures of the $(1,1,0)$ mode shape at 1.3 kHz and a sound pressure level of 145.4 dB. Results from the paint are compared with data from a Kulite pressure transducer and with linear acoustic theory. The paint may be used as a diagnostic technique for ultrasonic tests where high spatial resolution is essential, or in nonlinear acoustic applications such as shock tubes. [a] Currently at the Dept. of Aeronautics, U.S. Air Force Academy.

11:35

3aSP12. Performance of an infrasound source localization algorithm. Curt Szuberla, Kenneth Arnoult, and John Olson (Geophysical Inst., Univ. of Alaska Fairbanks, 903 Koyukuk Dr., Fairbanks, AK 99775-7320)

We present a performance analysis of a method of acoustic source localization based on time difference of arrival (TDOA) information for an arbitrary array of sensors. The method begins with the construction of a vector containing estimates of time delays for each unique sensor pair in the array via cross correlation. An optimization in the space of geographic location and sound speed is then conducted to minimize the difference between the observed vector and one calculated from a search in that space. The technique uses an analytic method of seeding based on the acoustic analog of a light cone. The source localization procedure is tested as a function of a number of parameters, including array aperture, number of sensors, sample rate, signal-to-noise ratio, and signal type. The technique is shown in application to real and synthetic infrasound signals and its statistical behavior is given.

Session 3aUWa

Underwater Acoustics and Acoustical Oceanography: Sediment Acoustic Processes: Quantifying the Effects of Sediment Properties on Dispersion, Attenuation, Reflection, Scattering, and Buried Target Detection III

Eric I. Thorsos, Cochair

Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6606

Masao Kimura, Cochair

Tokai Univ., Dept. of Geo-Environmental Technology, 3-20-1 Shimizu-Orido, Shizuoka 424-8610, Japan

Contributed Papers

7:45

3aUWa1. Seafloor manipulative experiments during sediment acoustics experiments (SAX99 and SAX04. Michael D. Richardson (Marine Geosciences Div., Naval Res. Lab., Stennis Space Center, MS 39529-5004) and Kevin L. Williams (Univ. of Washington, Seattle, WA 98105-6698)

Seafloor manipulations were performed during Sediment Acoustics Experiments (SAX99 and SAX04) in an attempt to test predictions of high-frequency acoustic backscattering models in sandy environments. Manipulations included alteration of seafloor roughness by raking to mimic the effects of sand ripples on acoustic scattering and placement of marbles, aluminum disks, and shells on the seafloor at different densities to mimic the effects of shelled animals and sand dollars on acoustic scattering (discrete surface scattering). In general, the predictions from scattering models and measured acoustic backscatter strengths were in agreement and the experimental results provide insight into natural phenomena. [Work supported by ONR.]

8:00

3aUWa2. A simple model for time-dependent scattering by changing seafloor roughness. Darrell R. Jackson, Kevin L. Williams, Christopher D. Jones (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, drj@apl.washington.edu), Michael D. Richardson, Kevin B. Briggs (Naval Res. Lab., Stennis Space Center, MS 39529-5004), and Anthony P. Lyons (Appl. Res. Lab., State College, PA 16804-0030)

The roughness of sandy seafloors is constantly changing due to biological and hydrodynamic activity. Even when these processes are in equilibrium so that the average acoustic backscattering intensity does not change, the character of successive backscatter echoes changes on time scales from minutes to days, with strong frequency dependence. This change can be characterized by the temporal correlation of the backscattered signal between consecutive pings, expressed as a lag-dependent correlation coefficient. A simple model for roughness change has been developed based upon the diffusion equation. This model assumes that newly created roughness is uncorrelated with existing roughness. It predicts exponential decay of the correlation with e-folding time $1/(DK_B^2)$, where K_B is the Bragg wave number, and D is the horizontal diffusivity. During SAX99 and SAX04, photographic and acoustic observations were made of changing natural and artificially generated roughness. The acoustic observations were made at 40 and 300 kHz, with the higher frequency exhibiting much more rapid decorrelation as predicted by the model. The diffusivities obtained by fitting the model to photographic and acoustic data ranged from roughly 10^{-9} to 10^{-10} m^{2}s^{-1} with the highest diffusivities found at locations where bottom-feeding fish were active. [Work funded by ONR.]

8:15

3aUWa3. Using forward scattering from a rippled sand/water interface to determine its flat surface reflection coefficient: Validation via simulation and use in testing sediment models. Kevin Williams (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, williams@apl.washington.edu)

Real-world sediments are inherently rough. At high frequencies (greater than 10 kHz) the roughness is typically sufficient to eliminate any coherent reflection at all but very shallow grazing angles. However, simulation results will be shown that demonstrate that an ensemble average of a large number $[O(50)]$ of forward-scattered signals from different regions of a rippled sand/water interface can be used to estimate the flat surface reflection coefficient for the interface. With this validation in hand, a forward scattering experiment carried out during a sediment acoustics experiment in 2004 (SAX04) will be described. The flat surface reflection coefficients derived from this experiment will be compared with predictions using a fluid and a Biot model for the sand sediment. This comparison indicates that a Biot sediment model is a more accurate predictor for the reflection coefficient. [Work supported by ONR.]

8:30

3aUWa4. High-resolution mapping and backscatter studies in support of the Ripples Program. Barbara Kraft, Christian de Moustier, Brian Calder, and Larry Mayer (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824, bjkraft@ccom.unh.edu)

The long-term goal of the Office of Naval Research's RIPPLES DRI is to develop the ability to predict the geometry and evolution of seabed morphology, particularly the presence or absence of ripples which have been shown to affect the sub-seabed penetration of acoustic energy. In support of the RIPPLES and SAX04 programs, high-resolution, multifrequency, multibeam mapping surveys were completed during September and October 2004 within three small areas off Destin, FL. The first survey was completed prior to Hurricane Ivan and followed the burial of minelike targets in the SAX04 field area. The second survey occurred during the SAX04 experiments approximately 6 weeks after Hurricane Ivan made landfall. Ripples were not observed in the bathymetric data from the first survey. Some small-scale features are present in the bathymetry from the second survey; however, most of the ripples generated by Hurricane Ivan had degraded by the time the survey was completed. Large-scale ripple features are present in the bathymetric data that extend off-shore from the SAX04 area. Difference maps of the pre- and post-Hurricane Ivan bathymetry and the acoustic backscatter data show remarkable evidence of sediment deposition and erosion. [Work supported by ONR Grant N00014-04-1-0624.]

3aUWa5. Doppler spreading of aircraft harmonics in a shallow-water channel off Kauai. David Barclay and Michael J. Buckingham (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA 02093-0238)

In September 2005, during the Makai experiments off Kauai, a light aircraft acted as a moving sound source over a shallow-ocean channel with a coral-sand bottom. Core samples yielded the porosity, bulk density, grain density, and grain size of the sediment. Acoustic sensors (a microphone in the air and a line of hydrophones in the isovelocity water column) recorded the sound of the aircraft as it flew over the sensor station. At least five engine/propeller harmonics, from 80 to 400 Hz, are present in the spectra of the hydrophone signals. Each harmonic exhibits Doppler spreading, with well-defined spectral peaks appearing on either side of the unshifted frequency. Those peaks with the greatest (up- and down-) shifts are associated with the shallow equivalent rays of normal modes in the water column. Two head-wave peaks, associated with slightly steeper rays, are less up- and down-shifted. Least shifted of all, because of their steep equivalent rays, are peaks due to modes supported by the sea surface and the subbottom, a few meters beneath the seabed. The spectral structure in the Doppler-spread harmonics is the basis of Doppler spectroscopy, an inversion technique for recovering the geoacoustic parameters of the seabed. [Research supported by ONR.]

3aUWa6. Using buried vector sensors to examine seabed layering in sandy sediment. Paul C. Hines, John C. Osler, Jeff Scrutton, David M.F. Chapman (Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, NS, Canada, B2Y 3Z7), and Anthony Lyons (The Penn State Univ., State College, PA 16804-0030)

There is considerable interest within the underwater acoustics community in quantifying the sound-speed dispersion in sandy sediments at low kiloHertz frequencies. To address this, data were collected during the SAX04 experiment (a US Office of Naval Research Departmental Research Initiative) using projectors and directional (vector) receivers buried in the top 1 m of the seabed, in conjunction with a water-borne source directly above the buried sensor field. Although using buried sensors significantly increased the complexity of the experimental setup, it should enable a direct time-of-flight measurement of acoustic wave speed, which simplifies interpretation of the results. However, initial analysis of the time-of-flight results indicates the existence of a buried layer, which increases the complexity of the data. In this paper, the vector sensors are used to steer beams to confirm the existence of the layer and to isolate the direct and reflected arrivals to improve the estimate of sound speed dispersion. [Work partially funded by US Office of Naval Research.]

3aUWa7. Estimation of received pulse from sea bottom with a transition layer using elastic finite difference time domain method. Takenobu Tsuchiya and Nobuyuki Endoh (Faculty of Eng., Kanagawa Univ., 3-27, Rokkakubashi, Kanagawa-ku, Yokohama, Japan)

Recently, ocean acoustic thermometry is being planned in shallow water. Characteristics of sound propagation in shallow water are influenced by bottom media because sound propagation is affected by acoustic boundary conditions of the water-sediment interface. To investigate characteristics of sound propagation in shallow water, it is necessary to develop an accurate numerical method to model sound propagation. Elastic finite difference time domain method (elastic FDTD) is very accurate and is useful for range-dependent models in shallow water. Therefore, we calculate the sound waveform reflected from the sea bottom with a transition layer in shallow water using elastic FDTD. It is assumed that a 1.5-kHz transducer is placed at 50-m depth. The influence of the sea bottom with the transition layer is calculated for a 400-m range. To determine the influence of thickness of the transition layer to sound propagation, we calculate sound pulses as transition layers of varying thickness. Results show that fluctuation of the pulse amplitude depends on the transition layer thickness if the thickness is greater than one wavelength. A sea bottom with a transition layer clearly affects sound propagation in shallow water.

3aUWa8. Stability of compressional waves in granular media. B. Edward McDonald (Naval Res. Lab., Washington, DC 20375)

Nonlinear compressional waves in a granular medium such as the seafloor are subject to the Hertzian nonlinearity of order 3/2 in the strain rate. As a result, the coefficient of quadratic nonlinearity becomes arbitrarily large at low stress. Two theoretical problems result from the Hertzian nonlinearity: (1) lowest order nonlinearity in the equation of state is no longer quadratic and (2) questions of stability arise when nonlinear steepening occurs arbitrarily fast. A variant of the NPE [McDonald and Kuperman, J. Acoust. Soc. Am. **81**, 1406–1417 (1987)] is used to derive a stability theorem for self-similar waves in a granular medium subject to three-dimensional perturbations. When the dominant nonlinearity is of order n between 1 and 3 (n is not necessarily an integer), the result suggests that wave stability is positively correlated with n. The method of characteristics is used to show that where the coefficient of nonlinearity diverges, the wave slope is forced to zero. [Work supported by the Office of Naval Research.]

3a THU. AM

Session 3aUWb

Underwater Acoustics: Sound Propagation: Measurements and Modeling I

Martin Siderius, Cochair
Heat Light & Sound Research Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130

Nobuyuki Endoh, Cochair
Kanagawa Univ., Dept. of Engineering, 3-27-1 Rokka Kubashi, Yokohama 221-8686, Japan

Chair's Introduction—9:55

Contributed Papers

10:00

3aUWb1. Benchmarking range-dependent seismo-acoustic propagation problems. Peter L. Nielsen, Finn B. Jensen, Mario Zampolli (NATO Undersea Res. Ctr., 19138 La Spezia, Italy), Michael D. Collins (Naval Res. Lab., Washington, DC 20375), and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

Developing the capability to accurately solve range-dependent problems in ocean acoustics has been going on for many years. A milestone in this development was the ASA benchmark effort in 1990 [J. Acoust. Soc. Am. **87**, 1497–1545 (1990)], where a set of models was applied to generic range-dependent acoustic fluid problems to establish a set of benchmark solutions for model verification purposes. A further benchmarking effort, which included shear waves in the bottom, was published in 1997 [IEEE J. Ocean. Engn. **22**, 226–236 (1997)] involving two seismo-acoustic codes based on spectral-integral techniques and a boundary-element code. Other codes have become available since then for solving general range-dependent elastic problems, either based on marching schemes that ignore backscattering (or adopt a single-scatter approximation) or techniques that solve the full boundary-value problem. In an attempt to establish a set of benchmark solutions for seismo-acoustic problems, we solve several low-frequency elastic problems involving changing bathymetry in multiple layers, and coupling between various wave types, including interface waves. The models in use or used are a parabolic equation model (RAM) and a commercial finite-element model (FEMLAB).

10:15

3aUWb2. Bathymetric and oceanic variability effects on acoustic propagation metrics in a shelf-break environment. Roger Oba (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

The interaction of oceanographic variability and bathymetry on acoustic propagation in the littoral is quite complex, and the relative importance is not readily quantified, especially for data-model comparisons. Sound speeds derived from ocean models could be compared directly to experimental sound-speed data, but this is a crude measure of environmental impact on acoustics. Instead, a 300-Hz source in a three-dimensional parabolic equation (PE) acoustic propagation model provides input to proposed acoustic metrics, especially related to array performance, to assess the effects of experimentally derived bathymetry and tide. The metrics can also be used for comparison of an ocean model and experimentally derived sound-speed profiles. Other acoustic mode metrics can be derived and compared to the more computationally intensive three-dimensional PE derived measures. As an example, the Navy Coastal Ocean Model (NCOM), a mesoscale oceanographic model, predicts sound-speed gridded on the order of 5 m in depth and 200 m in range at 6-h updates for the ASIAEx South China Sea site. This can then be used for metric calculations and data-model comparison. [This research is sponsored by the ONR.]

10:30

3aUWb3. Adaptive predictive-error optimal gridding. Donald R. DelBalzo and Lisa A. Pflug (Planning Systems Inc., 40201 Hwy. 190 E, Slidell, LA 70461)

Transmission loss computations in littoral areas often require a dense spatial and azimuthal grid to achieve acceptable accuracy and detail. The computational cost of accurate predictions in complex areas can be unacceptable for some Naval applications. EAGLE (Efficient Adaptive Gridder for Littoral Environments) was developed to reduce the computational requirements by adapting the fundamental calculation grid to the natural environmental structure. EAGLE produces a nonuniform grid structure within which the underlying data have approximately constant spatial uncertainty, or iso-deviance. The EAGLE point-selection approach is based on a predictive-error metric and least-squares polynomial curve fitting with arbitrary order. This research indicates the degree to which EAGLE grids are related to the underlying coherence properties of the natural environment. The results show significant efficiency enhancements compared to uniform grids of the same accuracy, when the environmental coherence properties are used to control the polynomial order. [Work sponsored by NAVAIR.]

10:45

3aUWb4. The effect of an ocean front on sound propagation in shallow water. Ying-Tsong Lin, Arthur E. Newhall, and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA, 02543, ytlin@whoi.edu)

From a perspective of sound propagation, a sloping bottom found in typical shallow-water environments can cause propagating sound to horizontally refract offshore. Additionally, an offshore ocean front can cause horizontal reflection/refraction shoreward. Combining these two effects, sound tends to propagate along the front, and "whispering gallery" modes can be observed. To model this effect, we first derived an analytic solution for the sound field using a point source in a rigid wedge environment that has a front placed offshore. The analytic solution gives us clear physical insight into this problem. For a more complicated environment, we exploited a 3-D acoustic propagation program in the ACOUSTICS TOOLBOX, which uses a vertical mode and horizontal Gaussian beam method. This model shows sound energy trapped ahead of the front with observable frequency and modal dependence. Also, when including the foot of the front, which is commonly seen on the continental shelf and extends inshore along the bottom, the model shows less modal attenuation, which raises the level of the trapped energy.

3aUWb5. Asymptotic expansions for normal modes in transparent wedge. Nick Maltsev (Guzuk Tech. Enterprises, 2443 Wyandotte St., Mountain View, CA 94043)

Asymptotic solutions of exact infinite system of equations for coefficients of angular Fourier harmonics of sound field have been built and computed for an arbitrary transparent wedge. Asymptotic solutions are compared with exact field in the case of two half-spaces and a rectangular wedge.

3aUWb6. Polynomial chaos expansions as extended Markov approximations in the propagation of uncertainty. Dennis Creamer (P.O. Box 660537, Arcadia, CA 91066)

The use of polynomial chaos for incorporating environmental variability into propagation models has been previously investigated where it was found that low-order expansions work quite well when the medium correlation length in the direction of propagation was small, i.e., the Markov approximation is roughly valid. It was suggested that an iterative scheme incorporating extensions of the Markov approximation might be useful. Using the method of successive approximations introduced by Klyatskin and Tatarskii (who used the method for investigating the validity of the Markov approximation), we show how the polynomial chaos equations can be written as a (smaller) set of equations that are basically nested expansions in the correlation function. This allows the use of a closure scheme, which does not involve truncating the expansion and which seems to be simpler to use. Comparisons are made between the approximation schemes in a geometrical optics limit.

3aUWb7. Polynomial chaos expansions for travel time and amplitude uncertainty in acoustic time series in ocean waveguides. Kevin D. LePage (Naval Res. Lab., Code 7144, 4555 Overlook Ave SW, Washington, DC 20375)

Polynomial chaos (PC) techniques have shown themselves to be a powerful technique for the estimation of the statistical properties of ocean acoustic propagation for situations where the vertical and horizontal structure of sound speed uncertainty can be specified in terms of empirical orthogonal functions (EOFs) with known amplitude probability distribution functions. Here we extend the CW solutions for the statistical properties of the normal models describing the acoustic field in the presence of internal waves with Gaussian distributed EOF descriptions to the time domain to describe the scintillation of the acoustic time series. Results are compared to exact solutions as well as to estimates based on the adiabatic approximation. Results show that the PC solutions do a much better job of capturing the true variability of the exact solutions for the time series variability than the adiabatic approximations does. [Work supported by ONR.]

3aUWb8. The propagation of acoustic field uncertainty in an ocean waveguide having finite correlation length sound speed variations. Steven Finette (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

While the physical laws governing wave propagation in ocean waveguides are well understood, the appropriate wave equations typically depend on parameters, fields, boundary, and initial conditions that are, at best, only partially known in any given environment. Incomplete environmental knowledge concerning these quantities is a form of uncertainty that should be characterized and quantified in any simulation-based prediction scheme involving the acoustic field. This talk focuses on the quantification of uncertainty in ocean waveguide propagation using a probabilistic formulation in which both the sound speed and acoustic fields are represented by stochastic basis expansions. An example is discussed for the case of an ocean waveguide with separable depth and range correlation functions describing the sound speed field. The relationship between this example and the propagation of acoustic field uncertainty in an internal wave field is discussed. [Work supported by ONR.]

3a THU. AM

Session 3pAAa

Architectural Acoustics and Noise: Acoustic Design for Green Buildings

Kenneth P. Roy, Cochair

Armstrong World Industries, Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17604

Tetsuya Sakuma, Cochair

Univ. of Tokyo, Graduate School of Frontier Science, Environmental Bldg. 627, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8563 Japan

Chair's Introduction—1:00

Invited Papers

1:05

3pAAa1. An overview of Leadership in Energy and Environmental Design (LEED) certification and acoustical design: Conflicts and synergies. Daniel C. Bruck (BRC Acoust. & Technol. Consulting, 1741 First Ave. South, Ste. 401, Seattle, WA 98134), David Lubman (14301 Middletown Ln., Westminster, CA 92683), and Louis C. Sutherland (27803 Longhill Dr., Palos Verdes, CA 90275)

The LEED program for green building certification acknowledges the importance of acoustics in building quality and design, but currently does not provide for specific credits in acoustics other than in the Innovation in Design category. This paper presents an overview of the LEED approach to sustainable design and discusses the design categories for LEED certification. Conflicts and synergies between the LEED system and acoustical design are discussed, including specific areas in which acoustics can contribute to sustainable design goals, such as recycled materials, energy efficiency, and interior environmental quality. Steps toward integrating acoustics into the LEED system are proposed, which incorporate elements of the Collaborative for High Performance Schools (CHPS) and ANSI S12.60, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools. The current direction of the U.S. Green Building Council with respect to incorporating acoustics into the LEED program is also presented.

1:25

3pAAa2. Acoustical lessons from four green buildings. Max Richter, Rosamund Hyde (Stantec, 1100-111 Dunsmuir St. Vancouver, BC V6B 6A3, Canada), and Murray Hodgson (University of British Columbia, V6T 1Z3, Canada)

Between 2002–2004, seven green buildings in British Columbia were studied during a Post Occupancy Evaluation (POE) research effort led by Stantec (formerly KEEN Engineering). Using the occupant satisfaction were survey developed by the Center for the Building Environment at UC Berkeley, data on eight aspects of occupant satisfaction were collected, including acoustical satisfaction. The BC buildings generally followed the trends that can be observed in green building acoustical performance, but none of the buildings scored as satisfactory. This year, Stantec and UBC measured four of the building in the original study to determine the acoustical characteristics of the building spaces. From this subsequent study, conclusions can be drawn about the relative acoustical performance of specific green strategies.

1:45

3pAAa3. Acoustical qualities of three Leadership in Energy and Environmental Design (LEED) certified buildings and why some disappoint. Steven D. Pettyjohn (The Acoust. & Vib. Group, 5700 Broadway, Sacramento, CA 95820-1852)

Obtaining points for acoustical qualities in office building or schools is difficult under the current LEED system. For this reason, the selection of partitions to provide speech privacy or prevent speech interference, the design of HVAC systems to meet background sound or vibration goals, or the inclusion of acoustical materials to provide the desired reverberation time is often either left out from the beginning or the first to be cut when budgets are tight. The results of sound tests done as a part of the commissioning of two office complexes and a new combination school, learning center, and administration building are shared. NIC tests were done on fixed and operable wall assemblies, background sound levels were measured in various spaces when due to mechanical and electrical systems and exterior sources, and reverberation time tests were done in conference rooms, classrooms, offices, and auditoria. The reason design goals were not met in two of the three complexes is related to budget cuts and a misunderstanding of the requirements and their importance to the successful use of the facilities. The commissioning process failed to correct the deficiencies.

<center>2:05</center>

3pAAa4. Acoustical evaluation of six "green" office buildings. Murray Hodgson (Acoust. & Noise Res. Group, SOEH-MECH, University of British Columbia, 3rd Fl., 2206 East Mall, Vancouver, BC V6T 1Z3, Canada), Rosamund Hyde, Blair Fulton, and Max Richter (Stantec Consulting Ltd., Vancouver, BC V6B 6A3, Canada)

As part of a project evaluating the performance of six "green" office buildings, detailed acoustical measurements were made to explain the positive and negative comments of building occupants obtained from a questionnaire survey. The objective was to determine how to design better green buildings. This paper presents an overview of the measurement protocol used, the measurements made, the results, and the lessons learned regarding optimal design.

<center>2:25</center>

3pAAa5. Acoustical evaluation of a "green" office building. Murray Hodgson (Acoust. & Noise Res. Group, SOEH-MECH, University of British Columbia, 3rd Fl., 2206 East Mall, Vancouver, BC V6T 1Z3 Canada), Rosamund Hyde, Blair Fulton, and Max Richter (Stantec Consulting Ltd., Vancouver, BC V6B 6A3, Canada)

As part of a project evaluating the performance of six "green" office buildings, detailed acoustical measurements were made on three floors of an office building in order to explain the positive and negative comments of building occupants obtained by a questionnaire survey. The objective was to determine how to design better green buildings. The building contained public spaces, private/closed offices, and open-plan-office areas. This paper presents details of the measurements performed, the results, and the design lessons learned.

<center>2:45</center>

3pAAa6. An overview of sound environmental aspect of Japan's comprehensive assessment system for building environmental efficiency (CASBEE) certifications. Tetsuya Sakuma and Fumiko Yano (Grad. School of Frontier Sci., The Univ. of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8563, Japan, sakuma@k.u-tokyo.ac.jp)

In the worldwide trend towards sustainable buildings, a new system called the Comprehensive Assessment System for Building Environmental Efficiency (CASBEE) has been established in Japan, as well as other existing assessment systems such as BREEM, LEED, GB Tool, etc. In particular, CASBEE is intended to evaluate building performance by a single-number index BEE, which is the ratio of environmental qualities to loadings. Concerning sound environment, three items are included in indoor quality evaluation: (1) noise (background noise, equipment noise); (2) insulation (window opening, wall, heavy/light floor impact sound); and (3) absorption; in addition, one item on noise emission outside the building site is included in environmental loading evaluation. However, there are many unclear points on the settings of items, ratings, and weightings, which will make trouble in practical application. In the last 3 years, several large cities have started a certification system using CASBEE for new large-scale buildings. In this presentation, an overview of the assessment system and the statistics of the city of Yokohama will be presented focusing on sound environmental aspect.

<center>3:05–3:25   Break</center>

<center>3:25</center>

3pAAa7. Achieving acoustical satisfaction in a green building. Ethan Salter, Charles M. Salter (Charles M. Salter Assoc., Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94104), Chuck Davis, and Scott Shell (EHDD Architecture, San Francisco, CA 94110)

The Carnegie Institution's Global Ecology Research Center at Stanford University has garnered many accolades, including the AIA's Excellence in Sustainability Award. This building incorporates many "green" and energy-saving design features, mechanical systems, and materials. The occupants of this facility have given it high marks in U.C. Berkeley's Center for the Built Environment's (CBE) survey of buildings. Staff at the Global Ecology Research Center are shown to be more satisfied with their acoustical environment than occupants of other green buildings surveyed by CBE. Measured acoustical data for speech privacy in open plan and enclosed conference rooms will be presented along with descriptions of acoustical design attributes for the building.

<center>3:45</center>

3pAAa8. Relationship between ventilation, air quality, and acoustics in "green" and "brown" buildings. Alireza Khaleghi, Karen Bartlett, and Murray Hodgson (SOEH-MECH, University of British Columbia, 3rd Fl., 2206 East Mall, Vancouver, BC V6T 1Z3, Canada)

This paper discusses a project involving direct monitoring of ventilation, indoor-air quality, and acoustical quality in "green" and "brown" buildings on the UBC campus. The objective was to determine the relationship between these various environmental factors, and the implications of the results for the ventilation-system concept/design, especially in green buildings. The buildings studied, the measurements performed, and the main results are described. The lessons learned about ventilation design are discussed.

4:05

3pAAa9. Acoustic and thermal evaluation low-cost dwelling, a case study in Curitiba, Brazil. Paulo H. Zannin (Mech. Eng. Dept., Federal Univ. of Parana, Rua Benedicto Berillo Fangueiro, 108, Santa Barbara, Curitiba, Parana, CEP, 81540-420, Brazil) and Eduardo Krueger (Cefet of Parana, Rebouas, Curitiba, PR)

The present paper presents acoustic and thermal performance evaluations carried out in low-cost houses located at the Technological Village of Curitiba. Two dwellings were evaluated, which were built according to different building systems: (1) the building system Batistella, built with prefabricated modules and elements, composed of reforestation wood, and (2) the building system Constroyer, composed of EPS plates, framed in a welded iron mesh, with structural plastering. For the thermal performance evaluation, it was verified that the materials used in the envelope were quite significant for controlling heat gains and losses through the building elements as well as for regulating heat storage. In this case, the system with highest thermal inertia presented better results. Results showed that both building systems had poor acoustic performance, characterized by a low weighted apparent sound reduction index of the facade $R'w = 18$ dB. The German DIN 4109 Norm recommends $R'w = 30$ dB. The weighted apparent sound reduction index measured for internal walls was $R'w = 25$ dB and $R'w = 30$ dB, for the building systems Battistela and Constroyer, respectively. The minimum required by German VDI 4100.

4:20

3pAAa10. Indoor occupant productivity perception and environmental sound measurement correlation. Brian J. Landsbeger, Liangcai Tan, and Xin Hu (Dept. of Mech. Eng., UNLV, 4505 Maryland Pkway, Las Vegas, NV 89154)

Occupant comfort and productivity have gained increasing attention by building owners and managers in comparison with traditional concerns about costs associated with building function and operating efficiency. To properly assess building performance, a comprehensive method for measurement and verification of integrated building performance was developed. Indoor sound, temperature, air movement, humidity, lighting, and air quality were measured for 3 days each in ten office buildings and ten schools. Concurrently, a perception questionnaire was completed by a subset of the occupants in the buildings and schools. The resulting questionnaire data were correlated with the physical measurements to determine what measures or combination of measures most accurately indicate worker perception of productivity variation when working under various environmental conditions. This presentation concentrates on correlation of acoustic measures, including dBA, NC, NCB, RC, and RC Mark II to worker perception of noise effects on productivity variation. Also, worker perception of productivity variation due to all measured indoor environmental factors was examined to determine additive effects and interaction between noise and other environmental factors. An ultimate goal of the study is to produce a method to quantitatively determine expected worker productivity variation based on measurable environmental factors.

THURSDAY AFTERNOON, 30 NOVEMBER 2006

OAHU ROOM, 1:00 TO 3:20 P.M.

Session 3pAAb

Architectural Acoustics: Sound Diffusion, Reflection, and Absorption

Jason E. Summers, Cochair
Naval Research Lab., 4555 Overlook Ave., S.W., Washington, DC 20375

Toru Otsuru, Cochair
Oita Univ., Faculty of Engineering, 700 Dannoharu, Oita 870-1192, Japan

Chair's Introduction—1:00

Contributed Papers

1:05

3pAAb1. Design of the diffuser with hemispheres using 1:10 and real scale models. Shin-ichi Sato and Jin Yong Jeon (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, s_sato@mac.com)

Sound diffusion by a wall structure is one of the main concerns with respect to the sound quality of concert halls. To quantify the sound diffusion of an actual hall from *in situ* measurements, the relationship between the scattering coefficient and the acoustical parameters measured in the laboratory should be investigated first. In this study, the scattering coefficient and acoustical parameters of the diffusers with hemisphere were investigated. The diameter, density, and the arrangement pattern of the hemisphere were investigated in terms of the scattering coefficient and the parameters obtained from the impulse response including temporal diffusion [H. Kuttruff, *Room Acoustics* (Elsevier Science, London, 1991)]. The results of measurements in 1:10 scale reverberation chamber with the same scale diffusers were verified by real scale results. Sound scattering index for reverberation chamber, which quantifies the scattered sound

field, was proposed. The optimum condition of the hemisphere diffuser specification maximizing the effect of scattering on a sound field was also investigated.

1:20

3pAAb2. Augmented-reality virtual-diffusion system: A tool to evaluate listener comfort. Pedro Novo (Acoust. Res. Ctr., Univ. of Salford, M5 4WT, UK, p.novo@salford.ac.uk)

The augmented-reality virtual-diffusion system presented here aims to evaluate the comfort of a subject listening to a nearby speaker located in a room filled with a diffuse crowd. The main parameter to be studied is sound field diffusion and its relation to comfort. To this purpose a virtual diffusion system, based on electro acoustic transducers mounted on the walls, is employed to actively change the properties of the diffuse sound field. The augmented reality factor is introduced through the interaction of the virtual diffusion system with the acoustics of the room. The present paper presents a brief description of the approach chosen to implement the augmented-reality virtual-diffusion system. Diffusion measurements obtained with both virtual and real diffusers are compared. Results regarding

auditory events location evoked by different sound fields and predictions from a cross-correlation-based binaural model employed to model these results are discussed. Initial results relative to comfort evoked by the different sound fields as well as a possible relation between auditory events location and listener comfort are discussed.

1:35

3pAAb3. Simulation of *in situ* measurement of diffusion and scattering coefficients. Diemer de Vries, Martin Kuster, Naziema Joeman, and Eric Schreurs (Lab. of Acoust. Imaging and Sound Control, Delft Univ. of Technol., P.O. Box 5046, 2600 GA Delft, The Netherlands)

Since it has become clear that simulation of the acoustical properties of a room is not successful unless also nonspecular reflections are taken into account, the determination of diffusion and scattering coefficients of boundary profiles is a hot topic in room acoustics research. Measurement of the impulse response of a boundary along a planar array of microphone positions enables us to calculate these coefficients *in situ*, avoiding the drawbacks of laboratory measurements (necessity of scale modeling, limited value of results for practical situations). Array technology makes it also possible to extrapolate the measured responses to the boundary surface, thus creating an acoustical image of the boundary. Next, scattering objects can be virtually added to this image and the impulse response of the modified configuration at the array position can be obtained by inverse extrapolation. This way, the effect of adding an object to a real environment is simulated in a highly realistic way. For several (virtual) modifications of the original boundary, diffusion and scattering coefficients have been calculated and compared. Also, listening tests have been carried out to assess the perceptual significance of the modifications.

1:50

3pAAb4. Effects of specular and diffuse reflections on perceived music quality. Pattra Smitthakorn, Martin A. Gold, Gary W. Siebein, and Bertram Y. Kinzey, Jr. (School of Architecture, Univ. of Florida, 134 ARCH, PO BOX 115702, Gainesville, FL 32611-5702, pattras@ufl.edu)

The purpose of this study was to investigate the effects of temporal distribution of diffuse and specular reflections on the perceived acoustic qualities of music performance. Sets of impulse responses were designed with different temporal distributions of early acoustic energy (specular and diffuse reflections). Then, three types of anechoic sound sources, orchestral music, trumpet, and piano, were convolved with the designed impulse responses. The results from the listening tests revealed that different room environments were needed to acoustically support different source characteristics: (1) A large number of diffuse reflections arriving within 40 and 80 ms of the direct sound improved perceived intimacy, texture, and overall impression for all sound sources; heightened perceived clarity for trumpet and piano; and reduced perceived glare for trumpet. (2) Diffuse reflections arriving between 80 and 160 ms of the direct sound preserved perceived reverberance and reduced perceived echoes as opposed to specular reflections arriving at the same time period. The results of this study indicate that music performance halls should be designed to include diffuse reflections from surfaces within the 80-ms time period to achieve preferred texture, intimacy, clarity, and overall impression and in the 160-ms time period to reduce echoes.

2:05

3pAAb5. The angular distribution of sound incident on a panel or opening. John L. Davy and Thomas K. John (RMIT Univ., GPO Box 2476V, Melbourne, Victoria 3001, Australia, john.davy@rmit.edu.au)

A theoretical method has been developed for predicting the directivity of the sound that is radiated from a panel or opening excited by sound incident on the other side. This directivity needs to be known when predicting the sound level at a particular position due to sound radiation from

a roof, wall, ventilating duct, or chimney flue. The method uses a two-dimensional strip model and the low-frequency result for a square piston. It was found necessary to use a weighting function in order to account for the angular distribution of the incident sound. Initially a cosine squared weighting function with a weighting angle parameter was chosen and the weighting angle parameter was varied in order to obtain the best agreement with experiment for the particular situation. This talk will describe the theoretical development of weighting functions, which are based on the actual physics of each situation. Situations that are covered include an opening or panel on the surface of a room and an opening at the end of a duct. The method will be compared with published experimental results on the directivity of the forced sound radiation from panels and the sound radiation from openings.

2:20

3pAAb6. Spatial coverage of reflector panels predicted with and without edge diffraction. Jonathan Rathsam, Lily M. Wang (Architectural Eng. Program, Univ. of Nebraska-Lincoln, 247 PKI, 1110 S. 67th St., Omaha, NE 68182-0681), and Rendell R. Torres (Rensselaer Polytechnic Inst., Troy, NY 12180-3590)

This investigation concerns the spatial coverage of reflector panels. The specular reflection zone (SRZ) is the spatial coverage predicted by geometrical acoustics, which is commonly used when designing reflector panels. A more accurate prediction of spatial coverage accounting for edge diffraction from the panels is called the effective scattered zone (ESZ). At higher frequencies, the SRZ and the ESZ are typically similar, but notable differences emerge at lower frequencies. In this study a variety of reflector panel arrays will be examined using a boundary element method (BEM) model based on the Kirchhoff-Helmholtz integral equation. The effective scattered zone is studied as a function of incidence angle, distance from array to source and receiver, and ka, the ratio of panel dimension to wavelength. The authors seek rules of thumb to predict configurations in which the SRZ and ESZ differ most notably. [Work supported by National Science Foundation.]

2:35

3pAAb7. Variable reflection acoustic wall system by active sound radiation. Genta Yamauchi and Akira Omoto (Omoto Lab., Dept. of Acoust. Design, Faculty of Design, Kyushu Univ., Shiobaru 4-9-1, Minami, Fukuoka 815-8540, Japan)

For acoustic design of special spaces such as studios, it might be necessary to realize a diffuse reflection from the specific wall surface. Moreover, this requirement often changes according to the time or situation. In this study, a new acoustic wall system is examined; it is a variable reflection acoustic wall system that can address such strict requirements. This system fundamentally comprises a loudspeaker, a pick-up microphone, and absorbent material. Varying the amplitude of the radiation sound that is picked up by the microphone and reradiated by the loudspeaker mounted behind the absorbing material therefore corresponds to varying the reflection performance of the wall system. Several models are proposed. The measurement results show that the proposed models can realize various absorption characteristics. Additionally, these models can realize various radiation characteristics: diffuse reflection and inclined reflection to a desired direction. Furthermore, the frequency characteristic of simulated radiation is changed, so the proposed system can generate reflection coefficient with intended frequency characteristics. The design of an effective feedback canceller between the pick-up microphones is the subject of ongoing research toward practical use of this proposed system.

2:50

3pAAb8. Helmholtz resonator array for low-frequency sound absorption. SangRyul Kim (Acoust. Lab., Korea Inst. of Machinery and Mater., 171 Jang-dong, Yuseong-gu, Daejeon, 305-343, Korea, srkim@kimm.re.kr) and Yang-Hann Kim (Korea Adv. Inst. of Sci. and Technol., Sci. Town, Daejeon, 305-701, Korea)

An array of Helmholtz resonators is often used for reducing low-frequency noise because of the high performance at its resonance frequency. One of the very attractive characteristics of the resonator array is that its effective frequency band is much wider than what can be obtained by a single resonator. This paper discusses a method of designing a Helmholtz resonator array panel for low-frequency sound absorption. First, various experimental results are introduced and studied. Second, based on the experimental results, a theoretical method is presented that can predict reasonably well the panel's absorption characteristics. Any configuration of resonator arrangement can be treated in the method, while all the resonators may or may not be identical. Comparisons between the predicted and experimental results reveal the accuracy of the proposed method. Finally, a numerical optimization is performed to design the Helmholtz resonator array panel. Some examples of optimal designs under some restrictions are also introduced.

3:05

3pAAb9. Concept of active reverberation box. Hisaharu Suzuki, Akira Omoto, and Kyoji Fujiwara (Kyushu Univ., Shiobaru 4-9-1, Minamiku, Fukuoka 815-8540, Japan)

A concept of a small reverberation box is proposed. This box has an active controller that can control the energy flow direction: it is a controller for the sound intensity. Consequently, the box is useful to measure the absorption characteristics of materials, even in a low-frequency range. The controller can create an arbitrary incident condition towards the materials. The measurement procedure with this "active reverberation box" is assumed as follows: (1) The incident angle of the sound is changed by the controller. (2) Impulse responses are measured in some points. (3) Steps 1 and 2 are repeated many times and impulse responses are averaged in an energy domain. (4) The reverberation time is calculated using the averaged impulse response. (5) Absorption characteristics are calculated by comparing the reverberation time of the empty room. The controller is formulized using a pseudo-Hermitian quadratic form of the sound intensity in the box. Numerical examination is performed using the finite difference time domain (FDTD) method. Experimental results demonstrate the effectiveness of the controller's performance. Availability of the active reverberation box is examined using some conventional measurement methods. Numerical results correspond approximately to the absorption coefficients that were taken in a reverberation chamber.

THURSDAY AFTERNOON, 30 NOVEMBER 2006 KOHALA/KONA ROOM, 1:00 TO 5:35 P.M.

Session 3pAB

Animal Bioacoustics: Vertebrate Vocal Production: Themes and Variations

W. Tecumseh Fitch, Cochair
Univ. of St. Andrews, School of Psychology, Fife KY16 9JP, U.K.

Kiyoshi Honda, Cochair
ATR Human Information and Processing Labs., 2-2 Hikaridai, Seika-Cho, Soraku-Gun, Kyoto 619-0288, Japan

Chair's Introduction—1:00

Invited Papers

1:05

3pAB1. Vertebrate vocal production: Themes and variations. W. Tecumseh Fitch (Univ. of St. Andrews, St. Andrews, Fife KY169JP, UK)

Most terrestrial vertebrates produce sound using a vocal system of lungs, larynx, and vocal tract, already in place in lungfish ancestral to all tetrapods. Despite the conservative nature of this basic system, there are a wide variety of modifications of the structure of the vocal system that appear to be linked to vocalization. In addition, several clades have evolved novel vocal production systems, including the avian syrinx and the odontocete nasal sac system. In this overview I will focus on the "standard" vertebrate system, discussing modifications of the vocal folds to produce high-pitched (e.g., bats and some primates) or low pitched (e.g., roaring cats) vocalizations, showing endoscopic videos of such unusual vocal folds in action. Moving to the vocal tract, a number of anatomical modifications result in vocal tract elongation, thus lowering formant frequencies. The flexible, dynamic nature of the mammalian vocal tract is illustrated with cineradiographs of vocalizing mammals. Finally, I will discuss the enigmatic existence of air sacs attached to the larynx or vocal tract. Vocal air sacs have evolved in parallel in many different species, in all classes. Despite a variety of plausible hypotheses, the function(s) of these air sacs remains poorly understood.

3pAB2. Simulation of Siberian tiger vocalization. Ingo R. Titze, Fariborz Alipour (The Univ. of Iowa, Iowa City, IA 52242; Natl. Ctr. for Voice and Speech, The Denver Ctr. for the Performing Arts, Denver, CO 80204), Wm. Tecumseh Fitch (Univ. of St. Andrews, St. Andrews, Fife, UK), Eric J. Hunter (The Denver Ctr. for the Performing Arts, Denver, CO 80204), and Edward J. Walsh (Creighton Univ., Omaha, NE)

The excised larynx of a Siberian tiger, which died of natural causes, was studied on a laboratory bench. Pressure-flow relations were determined, as well as modes of vibration of the vocal pads. In addition, CT scans of the animal were made available (from previous health examinations), which allowed the vocal-tract airway to be quantified. The larynx and vocal tract-data were then used to simulate the sound production of the animal. Results showed that multiple tissue modes are excited to produce a tiger roar, and that the production is highly efficient for aerodynamic to acoustic energy conversion. Scaling factors for linear dimensions of the larynx and vocal tract are derived to compare the phonation of this large mammal to human phonation.

3pAB3. Nonlinear phenomena in vertebrate vocal production. Hans-Peter Herzel, Riccardo Zaccarelli, Tobias Riede (Inst. for Theoretical Biol., Invalidenstr. 43, 10115 Berlin, Germany), Juergen Neubauer (UCLA, Los Angeles, CA 90095-1794), and W. Tecumseh FitchIII (Univ. of St. Andrews, KY16 9JU, Scotland, UK)

Nonlinear phenomena such as subharmonics, biphonation (two independent frequencies), and deterministic chaos are widespread in animal communication. Examples from primate vocalizations, African wild dogs, hyena whoops, and bird songs are presented. Despite the variability of size and morphology, vocal production in vertebrates is governed by the same underlying mechanism: vibrating tissues such as vocal folds, membranes, or labia are driven by overcritical airflow. Consequently, sophisticated biomechanical-aerodynamic models of human voice production can serve as a starting point for modeling nonhuman vocal production. However, there is a variety of additional anatomical features in diverse species: pronounced air sacs, extensions of the vocal folds termed ''vocal membranes'' or ''vocal lips,'' or the coordinated action of left and right syrinx in song birds. It is shown that such specific features can induce voice instabilities. For instance, source-tract interactions lead to subharmonics and chaos, and vocal membranes allow multiple ''registers.'' Finally bifurcations in biomechanical models of the syrinx are discussed. [Work supported by the Deutsche Forschungsgemeinschaft (Grant He 2168/7).]

3pAB4. Mosaic and multiple-step evolution of the anatomical foundations for human speech. Takeshi Nishimura (Dept. of Zoology, Kyoto Univ., Kitashirakawa Oiwake-cho, Sakyo, 606-8506 Kyoto, Japan, nishimur@anthro.zool.kyoto-u.ac.jp)

The human vocal tract develops to form a unique configuration with equally long horizontal and vertical cavities. This anatomy contributes greatly to the morphological foundations of human speech. It depends on the reduced growth of the facial length and on the major descent of the larynx relative to the palate. Anatomically, the laryngeal descent is accomplished through both the descent of the laryngeal skeleton relative to the hyoid and the descent of the hyoid relative to the palate. Here, the recent progresses are surveyed in the studies on the developments of the vocal tract anatomy in humans, chimpanzees, and macaques. They showed that humans and chimpanzees fully share the major descent of the larynx with the epiglottis descent. Macaques also show the hyoid descent, but this is not accompanied by the descent of the laryngeal skeleton and epiglottis. It is argued that the major descent of the larynx has probably evolved at least in two steps before the divergence of the human from the chimpanzee lineage, originally to confer an advantage via a function unrelated to speech. Thus, facial flattening has probably evolved only secondarily with anatomical consequences that facilitated speech in the human lineage. [Work supported by JSPS Grant-in-Aid (No. 16000326).]

3pAB5. Birdsong: The role of the vocal tract. Roderick A. Suthers (School of Medicine and Dept. of Biol., Jordan Hall, Indiana Univ., Bloomington, IN 47405, suthers@indiana.edu)

The avian vocal organ, the syrinx, is unique among vertebrates in its morphology and location at or near the base of the trachea. In oscine songbirds, two independent sound sources share the same vocal tract. Despite a growing knowledge of the neural and syringeal mechanisms underlying song, the acoustic role of the suprasyringeal vocal tract in modulating birdsong is still poorly understood. During song, the beak opening, or gape, is typically positively correlated with the dominant frequency of the vocalization, but the mechanism by which beak gape influences vocal tract acoustics is controversial. X-ray cinematography of singing birds reveals song-related movements of the hyoid apparatus and larynx that cause the volume of the upper vocal tract to vary inversely with the songs fundamental frequency. At low fundamental frequencies, the oropharynx and cranial end of the esophagus expand to form a single large chamber, the volume of which is reduced at higher frequencies by collapse of the esophagus and reduction of the oropharyngeal cavity. A computational acoustic model indicates that birds actively tune the volume of their oropharyngeal-esophageal cavity so that its major resonance tracks the fundamental frequency of the song. [Work supported by NIH.]

3pAB6. Odontocete nasalizations: Morphology, physiology, and evolution of sonar signal generation in toothed whales. Ted Cranford (Biol. Dept., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-0001)

The odontocete nasal apparatus contains tissue complexes that generate a variety of sounds. Sounds are produced by moving air across the phonic lips, internal structures that were formed by the invagination of the ancient nasal openings. X-ray computed tomography shows this sound generation complex bilaterally in all odontocetes except in sperm whales, where it is unilateral. Two bottlenose dolphins provided simultaneous bilateral recordings of internal nasal cavity pressure and nasal tissue motion. Small catheters measured pneumatic pressure changes within each bony nasal passage; high-speed video endoscopy recorded motion at both

3p THU. PM

pairs of phonic lips; and two hydrophones revealed acoustic pressure during target discrimination. Sonar pulses can be generated at the left and right phonic lips, independently or simultaneously. Air pressure in both bony nasal passages rises and falls together, even if the activity patterns at the two pairs of lips are different. Increasing pulse repetition rate or sound pressure levels are all normally accompanied by increasing nasal air pressure. The bilateral arrangement of sonar sources in the odontocete nose provides a basis for reinterpretation of previous findings; reevaluation of physiologic limits and on pulse repetition rate, signal bandwidth, frequency composition, and transmission beam geometry; and the evolution of this system.

Contributed Papers

3:05

3pAB7. The creative walrus: Novel sound production via contingency learning. Ronald J. Schusterman, Colleen Reichmuth Kastak (Inst. of Marine Sci., Long Marine Lab., Univ. of California, Santa Cruz, 100 Shaffer Rd., Santa Cruz, CA 95060), Debbie Quihuis (Six Flags Marine World, Vallejo, CA 94589), and Marla M. Holt (Univ. of California, Santa Cruz, Santa Cruz, CA, 95060)

Walruses (*Odobenus rosmarus*) have a range of anatomical specializations that can provide plasticity to their vocal emissions. These include respiratory adaptations related to breath-holding and buoyancy and soft tissue adaptations of the mouth, lips, and tongue related to suction feeding. The aim of this study was to examine the extent to which contingency learning using food reinforcement could modify the sound production of walruses. One male and three female captive walruses participated in the current study. Animals were first trained in air in a series of four experimental steps: (1) induce and reinforce natural sound production, (2) establish stimulus control over sound production and inhibition, (3) establish stimulus control over different sound types by mapping them to different discriminative stimuli, and (4) encourage the production of novel sounds by withholding reinforcement for previously emitted sounds and reinforcing only novel variants. Subsequently, two of the walruses were transferred from airborne to underwater testing. In the absence of additional training, both individuals, when prompted to vocalize, emitted a range of underwater sounds, many of which were qualitatively different from those produced in air and those recorded in baseline conditions under water. Thus, walruses, like cetaceans, can learn to modify their sound production.

3:20–3:35 Break

3:35

3pAB8. Acoustical development of the human baby cry: Anatomical and social factors. Yulri Nonaka, Noriko Kudo, Kazuo Okanoya (Lab. for Biolinguistics, RIKEN Brain Sci. Insitute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, ynonaka@brain.riken.jp), Noriko Kudo, Kazuo Okanoya (Chiba Univ.1-33 Yayoi-Cho, Inage, Chiba 263-8522, Japan), Noriko Kudo (JSPS), and Kazuo Okanoya (PRESTO, JST)

We investigated the acoustical development of the human baby cry based on our two-stage hypothesis, which holds that two distinct factors affect the development of the baby cry. First, humans possess the corticobulbar pathway connecting the face motor cortex and the medullary respiratory and vocal centers. This pathway is absent in nonhuman primates and other mammals. We postulate that the myelinization of this pathway would change the acoustic characteristics of the baby cry. Second, after neuroanatomical maturation enabled variable cry patterns, the social relationships with caregivers should differentiate the acoustical patterns of the baby cry. This should lead to family-specific cry patterns. To test the two-stage hypothesis, we recorded ten newborn babies until they were 1 year old. In all babies, the cry pattern was very simple up to 2 weeks, while after 1 month, the pattern became temporally and spectrally complex, suggesting neuroanatomical changes. At three months, we asked the mothers to classify the cries of their own babies into three categories: sleepy, hungry, and wet diaper. There were no acoustic patterns common to their classifications, suggesting that baby cry semantics are family-specific. These results support our two-stage hypothesis of cry development. [Work supported by JST and JSPS.]

3:50

3pAB9. Is the squirrel monkey larynx lateralized biomechanically? Charles H. Brown (Dept. of Psych., Univ. of South Alabama, Mobile, AL 36688, cbrown@usouthal.edu) and Fariborz Alipour (Univ. of Iowa, Iowa City, IA 52242)

Squirrel monkey larynges were dissected, mounted on a pseudotracheal tube, and phonated via the flow of compressed heated and humidified air. Synchronized audio and video signals were digitized and analyzed with a signal processing workstation. Image analysis algorithms measured the change in glottal area due to the movement of the left and right vocal folds, respectively, and measured the change in location of both vocal folds during phonation. The results show that the squirrel monkey larynx is capable of both unilateral and bilateral oscillation. In the unilateral case, oscillation is virtually confined to the left vocal fold. The converse of this has not been observed. In the case of bilateral oscillation, the folds may oscillate synchronously in phase, asynchronously out of phase, or at different fundamental frequencies. The results are consistent with the idea that the biomechanical properties of the left and right vocal folds differ with the left fold capable of oscillation at lower subglottal pressures and greater amplitudes. Changes in vocal fold elongation and adduction change the coupling between the two folds and produce different regimes of oscillation. Squirrel monkeys appear to employ intentional shifts between different regimes of phonation to expand the size of the vocal repertoire.

4:05

3pAB10. Measuring emotional arousal in the voiced sounds of two mammals, the rhesus monkey and African elephant. Joseph Soltis, Christina Wesolek, Anne Savage (Education and Sci., Disney's Animal Kingdom, Lake Buena Vista, FL), Kirsten Leong (Cornell Univ., Ithaca, NY), and John Newman (NICHD, Poolesville, MD)

Emotional arousal is expressed in the voiced sounds of primates and other mammals, but there are no consistent acoustic measures used and few comparative analyses. We apply a representative suite of source and filter measurements to rhesus macaque (*Macaca mulatta*) coo calls on Cayo Santiago, Puerto Rico, and African elephant (*Loxodonta africana*) rumble calls at Disney's Animal Kingdom, Florida, U.S.A. Based on social context, calls were classified into high and low arousal categories. In both species, MANOVA showed that the 15 measures taken together separated calls across arousal categories. In rhesus macaques, high arousal was associated with increased and more variable fundamental frequencies, increased amplitudes in the lower frequencies, and a shift in formant locations. In African elephants, increased arousal was associated with a shift of energy from lower to higher frequencies. In addition, low ranking females expressed a greater magnitude of acoustic change compared to high ranking females. The suite of acoustic features used here may successfully characterize arousal state in a variety of mammals, but the specific acoustic features that reflect arousal and the specific pattern of acoustic response may vary by species and individual. [Work supported by Grant Nos. NIH CM-5-P40RR003640-13 and NSF-IIS-0326395.]

4:20

3pAB11. Antiphonal exchanges in female African elephants. Katherine Leighty, Joseph Soltis, Anne Savage (Education and Sci., Disney's Animal Kingdom, Lake Buena Vista, FL), and Kirsten Leong (Cornell Univ., Ithaca, NY)

Female African elephants (Loxodonta africana) engage in antiphonal exchanges of rumble vocalizations. In this study, female African elephants ($N=7$) housed at Disney's Animal Kingdom were outfitted with audio-recording collars and videotaped during 50 1-hour observation sessions

(conducted in 2002). We found that production of the antiphonal response rumble reflected changes in herd affiliation that had taken place since our previous study (2000). Second, we examined response latency and found that females responded more quickly to the rumbles of affiliated partners than to those of a nonaffiliated female. Since affiliated partners spend more time in close proximity by definition, they may engage in higher rates of temporally associated calling simply because nearby stimuli impact them simultaneously or due to a social facilitation effect, not because they are true vocal exchanges. Therefore, we examined antiphonal rumble responses when female pairs were more than 25 m apart. We found that at these distances, affiliated females still exchanged rumbles more frequently, and responded at a decreased latency, than nonaffiliated pairs. Our results provide support for functional hypotheses of African elephant rumble vocalizations that require instances of temporally associated calling in order to be true communicative events. [Work supported by Grant No. NSF-IIS-0326395.]

4:35

3pAB12. Commonality in the vocalizations of order *Perissodactyla*. David G. Browning (Dept. of Phys., Univ. of Rhode Island, Kingston, RI 02881, decibeldb@aol.com) and Peter M. Scheifele (Univ. of Connecticut, Storrs, CT 06269)

As a follow-on to the human genome project, it has been suggested that the next major gene sequencing program may be for the horse as a representative of the relatively small order *Perissodactyla*. This order consists of three families; *Equidae* (equines), *Tapiridae* (tapirs), and *Rhinocerotidae* (rhinos), for a total of 17 species. Both sensual ability and environmental conditions (open plains to dense jungles) vary greatly among this small group; for example, horses have excellent eyesight, tapirs do not. Vocalizations of the various species are compared to determine any common characteristics that might serve as a basis for future association with specific genes. The universal quality appears to be the ability to vary pitch during the vocalization. Other subsets such as braying or whinnying exist among a limited number of species, which suggests that there may be evolutionary trends within the order.

4:50

3pAB13. Application of the source-filter theory to the study of loud-calling in male deer. David Reby (Dept. of Psych., Univ. of Sussex, Brighton BN2 9YU, UK)

I will review recent studies of roaring in red deer males, which illustrate how source-filter theory can be successfully applied to studying mammal vocal signals. Combining anatomical analyses of the vocal apparatus, acoustical analyses, and playback experiments using resynthesized calls, we have shown that (*i*) red deer and fallow deer males have a descended and mobile larynx, an anatomical innovation that was previously believed to be uniquely human and that enables callers to modulate their formants during vocalising; (*ii*) minimum formant frequencies provide an honest indication of body size in red deer roars, (*iii*) stags use rivals' minimum formant frequencies in assessment during male-male contests, and adjust the formants of their own replies in relation to what they hear; and (*iv*) hinds perceive size-related variation in formant frequencies, and pay more attention to roars in which formant spacing indicates larger

males, independently of fundamental frequency. I will then compare the acoustic structure of sexual loud calls in related subspecies (Corsican deer, Wapiti) and species (fallow deer, pere Davids deer), highlighting the variability of source- and filter-related acoustic features in this group. To conclude I will suggest directions for future research in this area.

5:05

3pAB14. Measuring amplitude in North American bison bellows: Environmental, social, and individual impacts. Megan Wyman, Lynette Hart, Brenda McCowan (Univ. of California-Davis, 1025 Haring Hall, One Shields Ave., Davis, CA 95616, mtwyman@ucdavis.edu), and Mike Mooring (Point Loma Nazarene Univ., San Diego, CA 92106)

Communication is a key aspect of any mating system and is a dramatic feature in the mating behavior of North American bison. During their reproductive season, bison bulls defend females from other males by establishing dominance through contests involving multiple signals, the most prominent being the bellow vocalization. While acknowledged that amplitude may be involved in sexual selection, it is rarely measured directly in the field (especially in large mammals) due to difficulties in accurately assessing this parameter. Using strict methodologies accounting for many sources of potential error, the impact that a variety of factors had on bellow amplitude was examined. Amplitude was measured in dB with $L_{a_{max}}$ and $L_{c_{peak}}$ weighting. Parameters *increasing* amplitude (order of decreasing impact): Number of surrounding bulls bellowing, wind levels, and number of surrounding bulls. Parameters *decreasing* amplitude (order of decreasing impact): Distance, head orientation, days spent in rut, time of day, number of copulations, dominance level of bull, and temperature. Since amplitude is correlated with measures of bull quality, condition, and surrounding social environments (beyond physical impacts of distance, etc.), it is likely that amplitude plays a significant role in competition over mates. [Work supported by NSF.]

5:20

3pAB15. Syllable chunking in Bengalese finches in a multi-tutor environment. Miki Takahasi, Iori Tomizawa, and Kazuo Okanoya (Lab. for Biolinguistics, RIKEN Brain Sci. Inst., 2-1 Hirosawa, Wako, Saitama, 351-0198 Japan, mtakahas@brain.riken.jp)

Experiments in birdsong learning have traditionally involved one tutor, usually the father, and the proportion and accuracy of note copying are discussed as indexes of song learning. But in the wild, there are several conspecific males around song-learning pupils. To test song learning under naturalistic situation, 11 male and ten female Bengalese finches were kept in a large flight cage containing ten pot-shaped nests. Seventy-eight chicks were born in this aviary over an 18-month period. We recorded the full songs of male birds after they had matured, and compared the songs between offspring and the original adult males. Most of them learned their songs from one or two tutors and some learned from three or four tutors. When a bird learned from multiple tutors, each of these tutor songs was copied in chunks and within a chunk the order of the song elements was the same as that observed in the tutor male. Furthermore, Bengalese finches change the order of chunks as they sing, and it was possible to express the rule using a finite-state syntax. Therefore, Bengalese finches have an innate universal grammar allowing them to construct a finite-state syntax from multiple sources. [Work supported by JSPS and JST.]

Session 3pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Celebration of Floyd Dunn II

Michael L. Oelze, Cochair

Univ. of Illinois, Electrical and Computer Engineering, Beckman Inst., 405 N. Mathews, Urbana, IL 61801

Shin-ichiro Umemura, Cochair

Kyoto Univ., School of Health Science, Kyoto 607-8507 Japan

Chair's Introduction—1:00

Invited Papers

1:05

3pBB1. A study on ultrasound parameter specification by temperature rise. Hiroshi Inoue and Chiaki Yamaya (Dept. of Electron. and Elec. Eng., Akita Univ., 1-1 Tegatagakuen-cho Akita-shi, Akita, 010-8502 Japan, inoueh@gipc.akita-u.ac.jp)

So far, temperature rises by absorption of ultrasound have been researched to secure the safety of the ultrasonic diagnostics and the application of ultrasonic therapy. This research is intended to develop a numerical simulation method that is applicable to estimation of medium constants, e.g., the absorption coefficient, using the temperature rise by the ultrasound. The temperature rises have been measured using a micro-thermocouple with high spatial resolution. Three-dimensional simulation using finite-difference time-domain (FDTD) method, which combines ultrasonic propagation and heat conduction, i.e., ultrasonic and temperature fields, is developed. At the first stage, two-dimensional analyses of ultrasonic fields were different from theoretical and experimental values. Therefore, we developed three-dimensional analysis. Time responses and spatial distributions of temperatures of the experiments and simulations were compared. Spatial distributions of temperature that were obtained using the FDTD simulation correspond qualitatively with those obtained by experiment, taking into account the thermocouple artifact. The possibility of estimation of medium constants is therefore suggested.

1:25

3pBB2. Modern acoustic microscopy and applications. Lawrence Kessler (Sonoscan, Inc., 2149 E. Pratt, Elk Grove Village, IL 60007)

The concept of acoustic microscopy was first suggested in the late 1930s by S. A. Sokolov and first demonstrated in a one-dimensional scan by Floyd Dunn and W. J. Fry in 1959. The notion of producing images of materials and biological tissues with UHF ultrasound to characterize their microelastic properties has broad appeal, however, most of the modern day (commercial) systems have been used instead for defect detection in manufactured products such as semiconductor devices. This paper will review a spectrum of applications from both the materials and biological worlds.

1:45

3pBB3. Twenty years' development of acoustic microscopy for medicine and biology. Yoshifumi Saijo, Hidehiko Sasaki, Motonao Tanaka (Dept of Medical Eng. & Cardiology, Tohoku Univ, 4-1 Seiryomachi, Aoba-ku, Sendai 980-8575, Japan, saijo@idac.tohoku.ac.jp), Naohiro Hozumi (Aichi Inst. of Technol., Yakusa-cho, Toyota 470-0392, Japan), Kazuto Kobayashi (Honda Electron. Co. Ltd., Oiwacho, Toyohashi 441-3193, Japan), and Floyd Dunn (Univ. of Illinois, Urbana-Champaign, IL)

Since 1985, scanning acoustic microscope (SAM) systems for medicine and biology has been developed and have been equipped for investigation of acoustic properties of various organs at Tohoku University. There are three objectives of SAM for biomedicine. First, it is useful for intraoperative pathological examination because staining is not required. Second, it helps understanding lower frequency ultrasonic images such as echography or intravascular ultrasound because SAM provides microscopic acoustic properties. Third, it provides information on biomechanics of tissues and cells. The most important feature of our SAM is to measure attenuation and sound speed of soft tissues by analyzing the frequency-dependent characteristics of amplitude and phase. Recently, a new-concept acoustic microscopy using a single pulsed wave instead of burst waves used in conventional SAM systems was proposed. Application of SAM has also been broadened and it covers not only myocardium and atherosclerosis in cardiology but also skin, bone, cartilage, tendon, and muscle in dermatology and orthopedic surgery. Higher resolution acoustic microscopy with over 500 MHz enabled us to visualize cultured cells and it can be applied for quality assessment of regenerated skins and cartilage. Impedance acoustic microscopy and 3-D acoustic microscopy are developed for *in vivo* tissue imaging.

3192   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3192**

3pBB4. VHF/UHF range bioultrasonic spectroscopy system and method. Naoyuki Akashi (Ichinoseki Natl. College of Technol., Ichinoseki 021-8511, Japan, akashi@ichinoseki.ac.jp), Mototaka Arakawa, Hiroyuki Odagawa, Toshio Sannomiya, Jun-ichi Kushibiki (Tohoku Univ., Sendai 980-8579, Japan), Noriyoshi Chubachi (Tohoku Univ., Sendai 980-8579, Japan), and Floyd Dunn (Univ. of Illinois, Urbana, IL 61801)

We developed the bioultrasonic spectroscopy system and method for doing ultrasonic tissue characterization studies in the VHF and UHF ranges. The system can determine the frequency characteristics of bulk acoustic properties of the sound velocity, attenuation, acoustic impedance, and density simultaneously by an ultrasonic transmission line comparison method using pure water as the reference, and using rf tone burst pulse signals. A biological tissue/liquid specimen is sandwiched between the parallel surfaces of buffer rods having ZnO piezoelectric film transducers on their opposite ends. Measurements are conducted with a typical thickness of 0.1 to 1 mm in the frequency range up to 700 MHz, and in the frequencies higher than 500 MHz, with a thin layer of specimen, several micrometers, operating in the multiple interference method. The system was applied for demonstration to measure the acoustic properties of bovine tissues, egg yolk and albumen, and aqueous solutions of biomacromolecules such as bovine hemoglobin and sugars in the frequency range 10 MHz to 1.6 GHz. The system and methodology will enable us to provide basic data for understanding the interaction of high-frequency ultrasonic waves with biological media and to make contributions to the establishment of the scanning acoustic microscope in the fields of biological and medical sciences.

3pBB5. Floyd's contribution to B/A measurement of biological materials. Leon A. Frizzell (Univ. of Illinois, 405 N. Mathews Ave., Urbana, IL 61801)

Ultrasound as used for medical diagnosis and high-intensity therapy is of sufficient amplitude that the effects of nonlinear propagation are important. Critical to modeling and understanding these nonlinear effects is knowledge of the nonlinearity parameter, B/A, for the biological tissues involved. In a series of papers Floyd Dunn and colleagues determined the value of B/A for several biological materials and soft tissues using the thermodynamic and finite amplitude methods. They showed that the B/A value depended directly upon the solute concentration in protein solutions. The value of B/A was around 7 for many soft tissues, including liver, muscle, and brain, but was around 11 for fat. Homogenized liver had a value that was less than that for whole liver, demonstrating an effect of structure. These results provided information regarding the range of B/A for biological materials and its dependence on concentration and structure and suggested the possible use of B/A in tissue characterization.

3pBB6. The impact of Floyd Dunn's research on echocardiography and myocardial tissue characterization. James G. Miller (Dept. of Phys., Washington Univ., St. Louis, MO 63130, james.g.miller@wustl.edu)

By the mid-1970s our lab had demonstrated that alterations in myocardium as a result of mature infarction could be quantified ultrasonically, with increased attenuation correlating with the extent of myocardial damage assessed in the same sites by depletion of creatine kinase. [Mimbs *et al.*, Circ. Res., **41**, 192–198 (1977)]. The Feigenbaum laboratory made the qualitative observation that zones of infarct were "more echo producing" than nearby zones. [Rasmussen *et al.*, Circulation **57**, 230–237 (1978)]. A 1973 article from Floyd Dunn's laboratory [S. Fields and F. Dunn, J. Acoust. Soc. Am. **54**, 809–812 (1973)] provided the basis for understanding these and many future developments in echocardiography and myocardial tissue characterization. In that seminal paper the authors state, "The magnitude of the echoes received is directly related to the preferred orientation and amount of collagen . . . ." We subsequently demonstrated the causal relationships between regional myocardial collagen content and both attenuation and backscatter. We also confirmed the Dunn prediction of anisotropy of attenuation, backscatter, and phase velocity. [Work supported by Grant NIH-R37-HL40302.]

3pBB7. Theoretical analysis of local attenuation approximations when calculating scatterer correlation length. Timothy A. Bigelow (Univ. of North Dakota, 243 Centennial Dr., Stop 7165, Grand Forks, ND 58202) and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

Determining the correlation length of ultrasound scatterers has shown potential in many tissue characterization applications. Accurate estimates of scatterer correlation length can only be obtained after correcting for focusing, frequency-dependent attenuation along the propagation path (total attenuation), and frequency-dependent attenuation in the region of interest (local attenuation). In order to correct for the local and total attenuation, an approximation is required because uncertainties in tissue attenuation can easily be 0.5 dB/cm-MHz, or greater, even for the same tissue type [J. Acoust. Soc. Am. **64**, 423 (1978)]. Earlier simulation results demonstrated that the most robust approximation for the local attenuation while completely compensating for total attenuation and focusing for weakly focused sources was to assume a finite frequency dependence of ∼0.8 dB/cm-MHz independent of the true attenuation of the tissue for attenuation values between 0.05 and 1 dB/cm-MHz. The goal of the present study was to investigate the theoretical basis for this approximation so that the results could be applied to other ranges of attenuation values. The theoretical analysis was then compared to the simulation results to validate the conclusions. [This work was supported by the University of Illinois Research Board and the University of North Dakota School of Engineering and Mines.]

3pBB8. High-frequency pulse-compression ultrasound imaging using an 18-MHz annular array. Jonathan Mamou, Jeffrey A. Ketterling, and Ronald H. Silverman (F. L. Lizzi Ctr. for Biomed. Eng., Riverside Res. Inst., 156 William St., New York, NY 10038)

High-frequency ultrasound (HFU) allows fine-resolution imaging at the expense of limited depth-of-field (DOF) and shallow acoustic penetration. Coded-excitation imaging allows a significant increase in the signal-to-noise ratio (SNR) and acoustic penetration depth. A 18-MHz, five-element annular array with a focal length of 31 mm and a total aperture of 10 mm was made using a 25-μm thick membrane. To excite the PVDF-TrFe transducer, an optimized 8-μs linear chirp spanning 6 to 30 MHz was uploaded into a waveform generator. The optimal chirp design was deduced from *ex vivo* experiments and computer simulations. Images of a seven-wire phantom were generated. The 25-μm wires were spaced in 1-mm axial and lateral increments. All 25 transmit/receive signal combinations were digitized and postprocessed for compression and synthetic focusing. Compression consisted of linearly filtering the signals with a mismatched filter. The mismatched filter was the time-reversed excitation chirp modulated by a Dolph-Chebyshev window. Marked image-quality improvements were demonstrated in images of *ex vivo* cow eyes. Results of this study demonstrate that the DOF can be improved by a factor of 3 and SNRs, after compression, can be improved by 24 dB.

Contributed Papers

3pBB9. Spatial distributions of acoustic parameters in high-intensity focused ultrasound fields. Vera A. Khokhlova (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, vera@apl.washington.edu), Olga S. Bessonova (Moscow State Univ., Moscow 119992, Russia), Michael S. Canney, Michael R. Bailey, and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

Different peak and average acoustic parameters determine the efficiency of different physical mechanisms of high-intensity focused ultrasound (HIFU) interaction with biological tissue. Spatial distributions of these parameters are therefore important for transducer calibration and extrapolation of measurements in water to application in tissue. In the case of linear focusing, all parameters of the acoustic field can be obtained from the spatial distribution of the wave amplitude. However, in nonlinear focused beams, each parameter has its own characteristic spatial structure, which changes with the increase of the HIFU power level. This work compares the focal size and location calculated for the peak positive and peak negative pressure, mean intensity, and effective acoustic energy absorption in water and in tissue. Numerical solutions, obtained with the KZK-type model, are analyzed for various regimes of linear, quasilinear, and strongly nonlinear propagation which includes formation of shocks. The results of simulations are validated by comparison with measurements performed with a fiber-optic probe hydrophone in water and in a tissue-mimicking phantom. The peak positive pressure and effective absorption are finely focused, whereas the negative pressure, responsible for cavitation, is broad and displaced towards the transducer. [Work supported by NSBRI SMS00402, RFBR, and NIH DK43881.]

3pBB11. Infinite periodic model of human lung. Yurii A. Ilinskii, Mark S. Wochner, Mark F. Hamilton, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

As described in a companion abstract, an investigation is underway to develop a bioacoustic model of human lung. In parallel with the numerical model described in the companion abstract, a semianalytic model is being developed. The basic unit cell, a truncated octahedron, is the same. In contrast here, the medium is assumed to be infinite in extent. The resulting lattice possesses cubic symmetry, and, for homogeneous deformation, all material properties are determined by a single unit cell. Deformation of the unit cell is determined by the 24 vertices of the polyhedron, only 6 of which are independent. In this presentation we discuss quasistatic deformation of the discretized medium, and thus ignore inertia and energy dissipation. An analytic model for the nonlinear elasticity of collagen and elastin is used to determine the stiffnesses of the springs connecting vertices of the polyhedra. Minimization of the potential energy for a given macroscopic deformation permits calculation of microscopic deformation within the unit cell, which in turn determines the stresses and therefore the elastic constants. Sample calculations will be presented, including the longitudinal and transverse wave speeds in the medium as functions of direction and orientation. [Work supported by ONR and ARL IR&D.]

3pBB10. Finite aperiodic model of human lung. Mark S. Wochner, Yurii A. Ilinskii, Mark F. Hamilton, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, mwochner@arlut.utexas.edu)

This work is focused on studying the dynamics of the human lung in response to low-frequency sound. The model developed uses individual alveoli as the basic unit cells, which are represented as truncated octahedra [Y. C. Fung, J. Appl. Physiol. **64**, 2132 (1988)]. Having 14 faces, this polyhedron possesses the smallest surface-to-volume ratio of any polyhedron capable of filling three-dimensional space. Unlike previous quasi-static models of this type, typically used to predict elastic properties of the lung, ours is a dynamical model. Mass is concentrated at the vertices of the polyhedra. Viscoelastic properties of the collagen and elastin in lung tissue are taken into account using a Kelvin model that is augmented to account for nonlinear elasticity. The Kelvin elements connect the point masses along the edges and across the faces of the polyhedra. Although not yet taken into account, the model permits inclusion of alveolar ducts by opening or closing selected faces of the polyhedra to accommodate airflow. Numerical simulations of small alveolar clusters subjected to different excitations and geometric constraints will be shown. [Work supported by ONR and ARL IR&D.]

3pBB12. Detecting spatially resolved inertial cavitation with a diagnostic ultrasound system. Caleb H. Farny, Robin O. Cleveland, R. Glynn Holt, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston MA 02215, cfarny@bu.edu)

The onset and presence of inertial cavitation in high-intensity focused ultrasound (HIFU) therapy has been identified as an important indicator of energy deposition for therapy planning. A common technique for detecting inertial cavitation is passive cavitation detection, where broadband noise emissions indicative of inertially driven bubbles collapses are detected using a fixed focus transducer. This technique is suboptimal for clinical applications, because most passive cavitation detector (PCD) transducers are tightly focused and thus afford limited spatial coverage of the HIFU focal region. Here, we report on efforts to use a diagnostic ultrasound system as a PCD by operating in passive mode and filtering out the main HIFU frequencies. This approach allows for spatially resolved detection of inertial cavitation throughout the focal region while being insensitive to scattering from voids or bubbles. [Work supported by the Dept. of the Army and the Center for Subsurface Sensing and Imaging Systems.]

5:00

3pBB13. Classification of flow-limiting thrombus in acute coronary syndrome by acoustic microscopy. Hidehiko Sasaki, Yoshifumi Saijo, Takayuki Kanno, and Motonao Tanaka (Dept of Medical Eng. & Cardiology, Tohoku Univ, 4-1 Seiryomachi, Aoba-ku, Sendai 980-8575, Japan)

Virtual histology (VH), based on a power spectral analysis of radio-frequency signal from intravascular ultrasound (IVUS) with neural network, can detect an unstable plaque from all the regions of coronary atherosclerosis. However, it provides limited information on plaque vulnerability because the resolution is larger than 100 m and thrombus is not classified precisely. The objective of the present study is to measure the acoustic properties of flow-limiting thrombus responsible for acute coronary syndrome (ACS) at the microscopic level. A newly developed acoustic microscope with 100-MHz pulsed ultrasound was equipped to measure two-dimensional distribution of attenuation and sound speed. The tissues were obtained from coronary arteries of 26 patients with ACS by intracoronary aspiration catheter. They were sectioned at 5-mm thickness and mounted onto glass slides without cover slips or staining and used acoustic microscopy study. The values of attenuation and sound speed of the red thrombus were significantly higher than those of the platelets-rich thrombus. Both acoustic parameters of old red thrombus were significantly lower than those of the organized thrombus. The acoustic properties provided by acoustic microscopy can classify the flow-limiting thrombus at different pathological stages. The criteria will provide further information for development of a VH-IVUS algorithm.

5:15

3pBB14. Image processing for scanning-type biological ultrasonic microscope considering its beam characteristics. Kazuto Kobayashi (Honda Electron. Co., Ltd. 20 Oyamazuka, Oiwa-cho, Toyohasi, Aichi, Japan), Naohiro Hozumi (Aichi Inst. of Technol., Yakusa Toyota Aichi Japan), and Shigeo Ohtsuki (Tokyo Inst. of Technol., Midori-ku, Yokohama, Kanagawa, Japan)

We have been developing a biological ultrasonic microscope with a high lateral resolution. An acoustic pulse is transmitted by a scanning transducer to a substrate on which a thin slice of a tissue is embedded. The transducer has a minute curvature in order to focus an acoustic beam with a high frequency component. It may be fabricated by MEMS technology; however, from the point of view of engineering, it may not be easy to guarantee the reproducibility of beam pattern as specified. In a high frequency range, a small error in fabrication will lead to a big difference in beam pattern, which strongly affects the observed image. Using newly proposed methodology to observe a well-defined pin-hole, we found that each commercial single probe transducer with the same specification has a unique beam pattern. In addition, the pattern was not always concentric. Consequently, we proposed a method to compensate the image that was obtained using an eccentric beam pattern. A clearer image was successfully obtained after the compensation.

THURSDAY AFTERNOON, 30 NOVEMBER 2006

NIIHAU ROOM, 1:00 TO 3:00 P.M.

Session 3pMU

Musical Acoustics: Musical and Structural Aspects of Instruments

Nicholas J. Giordano, Cochair
Purdue Univ., Dept. of Physics, West Lafayette, IN 47907

Isoharu Nishiguchi, Cochair
Kanagawa Inst. of Technology, 1030 Shimo-Ogino, Atsugi, Kanagawa 243-0292, Japan

Contributed Papers

1:00

3pMU1. Vibrations of a harp soundbox: Experiment and finite-element analysis. Chris Waltham and Andrzej Kotlicki (Dept. of Phys. & Astron., University of British Columbia, Vancouver BC, Canada V6T 1Z1, cew@phas.ubc.ca)

A medium-sized lever harp of proven design has been built. At each stage of construction, the components have been subjected to experimental modal analysis, using an impact hammer and small accelerometer, and also Chladni patterns. The results have been compared to a corresponding sequence of finite-element calculations. The aim is to build a picture of how the various ingredients—soundboard, veneer, ribs, soundbox, access holes, string tension, etc.—interact with each other to produce an acceptable instrument. Ultimately one would like to create a set of desiderata for the design and manufacture of harp components in a similar manner to that already achieved for, say, the violin.

1:15

3pMU2. The nonlinear motion of harp strings. Chris Waltham and Andrzej Kotlicki (Dept. of Phys. & Astron., University of British Columbia, Vancouver BC, Canada V6T 1Z1, cew@phas.ubc.ca)

The harp is an instrument with a set of plucked strings that excite the sound board directly, without the medium of a bridge. The strings are positioned at an acute angle to the plane of the sound board. The quality of the sound produced depends on the motion of the string and its interaction with the resonances of the sound board. String and sound board motions have been studied using small, infrared position sensors. The results are compared to those of a simple nonlinear model based on the measured elastic properties of the string materials and those of the soundboard. We will also describe the construction and calibration of the position sensors.

3195   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3195**

1:30

3pMU3. On Dagestan tamur acoustical features. Shamil Aliev
(DagDisel Factory, Lenina str. 1, Kaspiisk, Dagestan, Russia), Ernst
Myshinsky, and Alexander Pyshin (Krylov Inst., Saint-Petersburg,
196158, Russia)

The study of the tamur, a national Dagestan two-stringed musical in-
strument played by plucking with narrow body gouged from linden wood,
is presented. Distinctive features of the tamur sound making process are
described. The resonance properties of the tamur body, that is, the me-
chanical resonance system, are determined. Also the resonance properties
of air volume inside the tamur body, that is, an acoustic resonator, are
determined. An interaction of the tamur body mechanical vibration and air
volume inside the body oscillation are discussed. Tamur body sound ra-
diation dependencies on frequency are studied. The resonance properties
of instrument strings and purity of their sounding are determined. The
assessment of efficiency of transformation string oscillations into tamur
sounding board oscillations (body mechanical impedance) is discussed.
The elastic and acoustic constants of linden, that is, the wood material of
the tamur body made by gouge, are determined. The logarithmic decre-
ment of tamur sounding board oscillations is determined. The typical spec-
tra of tamur sounding are analyzed.

1:45

**3pMU4. Comparing harmonic reinforcement in the Octet's Mezzo
and a typical Strad-model violin.** Duane Voskuil (Univ. of North
Dakota, Bismarck, ND, dvoskuil@bis.midco.net) and Carleen M.
Hutchins (The New Violin Family Assoc., Wolfeboro, NH)

Many players and listeners maintain the Octet violins have a fuller
sound than conventional instruments. Preliminary data indicate Octet vio-
lins may better embody some resonant intervals that make this more likely.
A theory is presented based on a study comparing the Octet's Mezzo and
a Strad-model violin, concluding that C. M. Hutchin's original design
principle for the Mezzo fortuitously related A0 to A1 as an interval of a
fifth near the two open middle strings, allowing more constructive wave
reinforcement than conventional violins. The theory considers other inter-
vals, seeking to maximize harmonic reinforcement by modifying instru-
ments, (1) so that every bowed note is reinforced by an activated wood or
air resonance by placing the notes fundamental and overtones close
enough to a resonant frequency to activate it, (2) so that two resonances
constructively interfere when they are simultaneously activated because
they are separated by a harmonic interval, and (3) so that as many higher
mode resonances as possible are activated by upper partials of the bowed
notes harmonic series.

2:00

3pMU5. Linearized and nonlinear dynamics of bowed bars. Octavio
Inacio (Escola Sup. de Musica e Artes do Espectaculo, Musical Acoust.
Lab., R. da Alegria, 503, 4000-045 Porto, Portugal,
octavioinacio@esmae-ipp.pt) and Jose Antunes (Instituto Tecnologico e
Nuclear, 2686 Sacavem codex, Portugal)

Friction-excited instruments have been for many decades an inexhaust-
ible source of physical delight. In the last decades, spectacular improve-
ments in computational power and in numerical methods enabled simula-
tions of the self-excited nonlinear regimes with considerable realism and
detail. The authors of the present paper have achieved many such nonlin-
ear simulations, using a powerful modal approach, and decided to inves-
tigate here in detail the characteristics of the linearized and nonlinear
regimes of bowed bars. After stating the theoretical modeling approach for
the nonlinear problem, we derive a corresponding linearized model, from
which the complex eigenvalues and eigenvectors are computed as a func-
tion of the bowing parameters (friction parameters, normal force, and tan-
gential velocity, accounting for the bowing location). We thus obtain plots
of the modal frequencies, damping values, and complex mode shapes, as a

function of the bowing parameters, as well as stability charts for each one
of the system modes. When compared with the nonlinear motion regimes,
these results offer interesting information concerning the stability behavior
of the system, and further insight when addressing the postinstability non-
linear limit-cycle responses. Several differently tuned bars are addressed
in order to assert the configurations leading to optimal playability.

2:15

**3pMU6. Spectral method applied in physical modeling synthesis of
plate.** Lin Zhang (School of Phys., Univ. of Edinburgh, Kings Bldg.,
Edinburgh, EH9 3JZ, UK, lin.zhang@ed.ac.uk)

As a well-applied method in scientific computation, spectral methods
have properties like higher order accuracy, nonlinearity expandability, and
memory saving. By an example, this work will explore the potential of its
application into physical-based sound synthesis With background work of
finite difference on wave equation, spectral differentiation is further ap-
plied in space of discretization of partial difference equation (PDE). Gen-
eral plates (including membrane) percussion models are created numeri-
cally presenting the wave propagation inside of plate. According to two
different dimensions (rectangular and circular), Fourier or Chebyshev in-
terpolations are applied. In addition, stiff system is introduced by adding
sizzlelike partials onto plate model (sizzled cymbals in jazz music) and
complex collision phenomena (for example, ride cymbals). Problems such
as stability condition take an important position in analysis. The algorithm
is to generate high-quality multichannel audio outputs for musical pur-
pose. For the continuity and expandability as application module, parallel
computation programming is taken as the focus of final implementation on
a Sun Bluefire E15K shared memory machine [Work supported by EPCC.]

2:30

3pMU7. Vibrational characteristics of the glassharp. Koichi Uchida
and Kenshi Kishi (Dept. of Electron. Eng., Univ. of Electro-Commun.,
1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585 Japan
coici@crane.ee.uec.ac.jp)

Using the finite-element method (FEM), glassharp vibration was ana-
lyzed. Results show that the glassharp pitch is controllable at the stage of
manufacturing. First, the effects of several factors were studied, such as
geometry approximation, element type, and the number of division on the
accuracy of mode frequencies. Second, it was examined how each mode
frequency changes by the fixed position and length of the glassharp stem.
The results demonstrate that their influences were of no great importance.
Third, the pitch was widely adjustable by changing the swell of the cup.
Particularly, it was finely adjustable by cutting a small circular groove
inside a bottom of the cup instead of a conventional method.

2:45

**3pMU8. Physical modeling of Tibetan singing bowls using dynamic
forcing.** David M. Etlinger (School of Music, Northwestern Univ., 711
Elgin Rd., Evanston, IL 60208)

The geometric simplicity of Tibetan singing bowls produces surpris-
ingly complex vibrational behavior. The curvature of the bowls causes
strong coupling between modes, and the physical parameters of actual
bowls (namely their thickness and deep curvature, as well as their asym-
metrical excitation) produce higher-order effects that cannot be neglected.
In this paper we present a physical model for the bowl's acoustical output.
The model is based on analytical solutions for the natural modes, and
predicts both transient and steady-state behavior. Using the method of
dynamic forcing, we investigate bowl response for linear (striking) and
nonlinear (rubbing) excitation methods, producing audio as well as nu-
merical data. Finally, we use the model to gain insight into aesthetics and
design: by comparing our simulations to actual antique or high-quality
bowls, we can begin to determine what parameters contribute most to
creating the desired sound.

Session 3pNSa

Noise, Speech Communication, and ASA Committee on Standards: Progress in Hearing Protection and Hearing Conservation II

Elliott H. Berger, Cochair
E-A-R Aearo Co., 7911 Zionsville Rd., Indianapolis, IN 46268-1657

Hiro Takinami, Cochair
RION Co., Ltd., 3-20-41 Higashimotomachi, Kokubunji, Tokyo 185-8533, Japan

Contributed Papers

1:00

3pNSa1. Field study of sound exposure by personal stereo. Rodrigo Ordoñez, Karen Reuter, and Dorte Hammershøi (Dept. of Acoust., Aalborg Univ., Fredrik Bajers Vej 7-B5, DK-9220 Aalborg Ø, Denmark, rop@acoustics.aau.dk)

A number of large scale studies suggest that the exposure level used with personal stereo systems should raise concern. High levels can be produced by most commercially available mp3 players, and they are generally used in high background noise levels (i.e., while in a bus or train). A field study on young people's habitual sound exposure to personal stereos has been carried out using a measurement method according to principles of ISO 11904-2:2004. Additionally the state of their hearing has also been assessed. This presentation deals with the methodological aspects relating to the quantification of habitual use, estimation of listening levels and exposure levels, and assessment of their state of hearing, by either threshold determination or OAE measurement, with a special view to the general validity of the results (uncertainty factors and their magnitude).

1:15

3pNSa2. Antioxidant use in the prevention of noise-induced hearing loss and ototoxic chemicals in the workplace. Sandra C. MacLean Uberuaga (Alaska Occupational Audiol. & Health Services, 200 West 24th Ave., #196, Anchorage, AK 99503)

There have been discoveries in auditory research as to how the cochlea is injured during noise exposure and how it may sometimes be possible to reverse or prevent acute noise damage with medications. There are two primary mechanisms of damage; mechanical, as delicate tissue structures are torn apart when loud blasts occur above 125–130 dB SPL, and metabolic exhaustion, which commonly results from habitual noise exposure. During metabolic exhaustion toxic waste products (free radicals) form as cells in the cochlea are stressed. Our bodies can react to noise trauma by presenting a defense of antioxidant enzymes. However, when these defenses are overwhelmed, the hair cell is subject to serious damage to its

nuclear DNA, mitochondria, and membranes. This research has been promising in animal studies. We will discuss pharmacological agents that can be used by the ear to synthesize the toxic products. The next step is well-designed human clinical trials in occupational settings where extreme noise levels cause hearing loss despite the seemingly appropriate use of hearing protection devices (e.g., aircraft carrier decks, mining and petroleum industries). Chemicals and agents that can cause hearing loss in the workplace will also be reviewed.

1:30

3pNSa3. Combined effects of occupational exposure to noise and organic solvents on the auditory system. Ikuharu Morioka (School of Health and Nursing Sci., Wakayama Medical Univ., 580 Mikazura, Wakayama, 641-0011 Japan), Yoshihiro Minami, Nobuyuki Miyai, Kazuhisa Miyashita (Wakayama Medical Univ., Wakayama, Japan), Hiroichi Yamamoto (Osaka Occupational Health Service Ctr., Osaka, Japan), and Orawan Kaewboonchoo (Mahidol Univ., Bangkok, Thailand)

Noise exposure is the most common cause of hearing loss. In many work environments where organic solvent exposure occurs, a high level of noise exposure is also common. However, as to the ototoxic effects of organic solvents, the earlier studies focused on the exposure to an organic solvent alone. It is necessary to clarify the combined effects of exposure to noise and organic solvents at the simulated situation of the work environments. The objective of this study was to review the literature on the ototoxic effects of exposure to organic solvents combined with noise. Two organic solvents were selected: methanol, a widely used organic solvents in the occupational field, and styrene, also widely used and a human neurotoxicant. Judging from the results of animal experiments and epidemiologic studies, a synergistical interaction of noise with such organic solvents' ototoxicity was clarified even at the lower concentration where the organic solvent alone showed no ototoxicity in the experimental animals and at the concentration lower than the occupational exposure limits of organic solvents. Consequently, it is important to carefully access the occupational exposure limits when workers are exposed to organic solvents combined with noise, and to take measures against combined exposure.

3p THU. PM

Session 3pNSb

Noise: Active Noise Control

Yukio Iwaya, Chair

Tohoku Univ., Research Inst. of Electrical Communication, 2-1-1 Katahira, Aoba-Ku, Sendai 980-8577, Japan

Chair's Introduction—1:55

Contributed Papers

2:00

3pNSb1. Active noise control-like performance of a spherical reflector.
Kunisato Seto, Md. Tawhidul Islam Khan, Kenbu Teramoto (Dept. Mech. Eng., Saga Univ. 1 Honjo-machi, Saga-shi, Saga 840-8502, Japan, kunisato@cc.saga-u.ac.jp), and Hiromitsu Ohta (Natl. Fishery Univ. Shimonoseki 759-6595, Japan)

The principle of cancellation of tonal noise by using a spherical reflector, the designing method of an appropriate reflector for canceling jet screech, the advantage of a spherical compared to a flat reflector for canceling a tonal noise, and a prospective application for an aircraft to protect the fuselage from acoustic fatigue are discussed. The thrust loss due to the new technique was also evaluated directly with a vertical wind tunnel. The change in the aerodynamic structure was optically observed by using a schlieren apparatus along with a high-speed video camera. [Work supported by grant of JSPS, Japan.]

2:15

3pNSb2. A study of subjective effect on active noise control system.
Sayoko Utsumi, Eiji Murai, Yuko Watanabe, and Hareo Hamada (SIE, Tokyo Denki Univ., 2-1200 Muzai-gakuendai, Inzai, Chiba, 270-1382, Japan)

In recent years, the influence of various noises on human beings in our daily life has become a problem. Especially, the low-frequency noise has been considered to have a serious influence on work efficiency of people who are working under the noise exposure. On the other hand, active noise control (ANC) systems have achieved the efficient reduction noise, especially for the low-frequency noise. Actually, a number of successful ANC systems have been reported. Therefore, ANC systems may improve our work efficiency. In this paper, features of various noises are observed to get a representative value to be a sound source for subjective listening tests. Then, the effectiveness of the ANC system is evaluated to observe a physiological reaction of subjects. For that, response times to the short-term memory task will be measured in three different conditions, such as usual condition, under the noise exposure, and that with a running ANC system.

2:30

3pNSb3. Near-field mapping of pressure fields during active noise control of small axial cooling fans. Benjamin M. Shafer, Kent L. Gee, Scott D. Sommerfeldt, and Cole V. Duke (Brigham Young Univ., Provo, UT 84602)

In the past, tonal noise from small axial fans has been globally reduced using active noise control (ANC) with near-field error sensors placed according to a theoretical condition of minimized radiated power [K. L. Gee and S. D. Sommerfeldt, J. Acoust. Soc. Am. **115**, 228–236 (2004)]. The theoretical model, based on mutual coupling of point sources, showed that pressure nulls exist in the near field when the radiated power is minimized. Error sensor placement at these locations should then optimize global ANC. This study comprises an experimental investigation in which the actual locations of these near-field pressure nulls have been measured over a two-dimensional grid with a linear array of microphones. The array consists of 25 quarter-inch microphones with half-inch spacing. This array has been used to map the pressure field from a 60-mm cooling fan during ANC, in addition to a benchmark case, where a small loudspeaker has been mounted in place of the fan. The experimental results are compared to the theoretical pressure null locations in order to determine the efficacy of the point source theoretical model.

2:45

3pNSb4. Active control of free-field sound using near-field energy-based error signals. Ryan T. Chester and Timothy W. Leishman (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

Practical efforts to actively control sound often require error sensors located in the acoustic or geometric near field of primary and secondary sources. Unfortunately, when using conventional acoustic pressure sensors, control performance often becomes very sensitive to sensor position, and optimal sensor positions are difficult to ascertain or maintain with changing system conditions. Other types of error sensors or combinations of error sensors may produce fewer position-dependent complications, yielding excellent global sound-field control with greater ease and consistency. This paper compares the use of potential, kinetic, and total energy density error signals in the near field of a primary and secondary source in free space. It also discusses the distinct spatial and spectral uniformities of these quantities and their impact on the active control of sound.

Session 3pNSc

Noise, Architectural Acoustics, and Speech Communication: Acoustical Design Factors in Spaces Where Speech Communication is Critical

Todd R. Beiler, Chair
343 Lapa Pl., Kailua, HI 96734

Chair's Introduction—3:10

Contributed Papers

3:15

3pNSc1. Can operating room nurses accurately classify noise exposures? Bernadette Stringer and Ted Haines (Dept. of Clinical Epidemiology and Biostatistics, McMaster Univ., ON, Canada)

Worker's qualitative chemical exposure estimates when compared to estimates made by experts or with monitoring devices have been found to be accurate. But almost no research exists on worker's ability to classify noise exposure. This investigation took place within a multi-hospital intervention study in which OR nurses completed a questionnaire at the end of each surgery and answered, "During this surgery could you easily hear: quiet talking, normal talking, or loud talking?" In 255 surgeries, noise was measured using a sound level meter, for a minimum of 15 min, while in 68 surgeries noise was measured for >70% of the surgery. In the 255 surgeries in which noise was measured for 15 min or >, it was found that nurses who heard quiet talking were exposed to 62.8 dB(A) and that nurses who heard normal or loud talking were exposed to 65.1 dB(A), a difference that was statistically significant ($p = 0.019$). In the 68 surgeries where noise measurements lasted more than 70% of the surgery, nurses who heard quiet talking were exposed to 64.0 dB(A), while nurses who heard normal or loud talking were exposed to 67.3 dB(A), a difference that was borderline significant ($p = 0.07$). Nurses can distinguish between noise that interferes with quiet talking during surgery and noise that does not. [Work funded by Ontarios Workplace Safety Insurance Board.]

3:30

3pNSc2. Noise in Canadian hospital operating rooms. Ted Haines and Bernadette Stringer (Dept. of Clinical Epidemiology and Biostatistics, McMaster Univ., ON, Canada)

Noise levels in hospital operating rooms have not been well characterized. Therefore, noise levels were measured using a sound level meter in a sample of surgeries included in a multihospital intervention study, assessing the effectiveness of a recommended work practice to decrease occupational exposure to blood during surgery. The duration of the measurements ranged from 15 min to several hours. Among types of surgery for which at least four measurements were done, the L_{eq} for orthopedic surgery was the highest at 70.1 dB(A) (range 60.8–75.1), followed by 63.7 dB(A) for neurosurgery (range 57.4–68.1), and 62.8 dB(A) (range 58.5–70.3), for general surgery. Gynecological surgery had the lowest L_{eq}, 60.8 dB(A) (range 56.5–62.5). Peak levels were found to be as high as 132.8 and 132.6 dB(A), in general and orthopedic surgery, respectively, and lowest in neurosurgery at 102.6 dB(A). These noise levels are consistent with those from a comprehensive U.S. study, and substantially exceed EPAs recommended level of 45 dB(A) for hospitals. [Work funded by Ontarios Workplace Safety Insurance Board.]

3:45

3pNSc3. Noise levels with the potential to impede communication during surgery. John Oudyk and Bernadette Stringer (Occupational Health Clinics for ON Workers Inc., 848 Main St. E., Hamilton, ON L8M 1L9, Canada)

Surgical teams consist of personnel who may or may not regularly work together, among whom communication can be more difficult because of surgical masks, face shields, or goggles, factors believed to have contributed to a recent serious medication error that occurred when the anaesthetist misunderstood the surgeon's request. In a multihospital, occupational health intervention study a sound level meter was used to measure ambient noise for 15 min or > during 256 surgeries included in the study. Results revealed an L_{eq} of 70.7 dB(A) and peak of 115.8 dB(A) (at the 95th percentile). Since the clarity of words depends on maintaining a signal-to-noise ratio of at least 15 dB(A), personnel working with such noise would likely increase their speaking levels up to 75–85 dB(A), a level substantially higher than normal speech ranging from 55–65 dB(A). Although the effect of noise levels on patient safety, or the health of OR personnel who must raise their voices to be heard, as well as on the overall climate that results when noise is elevated, have not been thoroughly investigated, evidence of negative effects is growing. And, while further characterization of OR noise is justified, this should be carried out to develop tailored interventions that can then be evaluated. [Work was funded by Ontario's Workplace Safety Insurance Board.]

4:00

3pNSc4. Comparison of speech communication of renovated classroom by subjective method. Coriolanus Lam, Chun Wah Leung (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hung Hom, Hong Kong, coriolanus.lam@polyu.edu.hk.), and Kai Ming Li (Purdue Univ., IN)

This paper presents measurement results on the acoustical characteristics of two standard and identical classrooms, one of which has been renovated. Several acoustical quantities that affect speech communication have been studied. These quantities were used to predict and quantify the characteristics and the qualities of measured classrooms that focused on speech communication. Subjective method by speech intelligibility tests was conducted in the measurements. Phonetically balanced monosyllabic word tests in both English and Chinese were carried out. In addition, public address messages, both in English and Chinese, from Hong Kong Mass Transit Railway were used as material for the tests. The results indicate that the renovation was beneficial on the quality of speech communication in general except for echo effects and increase in reverberation under several combinations of acoustical settings. These results were used to evaluate the relationship between speech perception quality and speech intelligibility. Besides, it will be used for new classroom designs in Hong Kong for better speech communication. [Work supported by Research

3p THU. PM

4:15

3pNSc5. Case study: Predict acoustical quality of renovated classroom. Coriolanus Lam, Chun Wah Leung (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hung Hom, Hong Kong, coriolanus.lam@polyu.edu.hk), and Kai Ming Li (Purdue Univ., IN)

The study on prediction of acoustical quality of two standard and identical classrooms, one of which has been renovated, was presented. Acoustical back-wall panels and sound-reflecting panels were installed in the renovated classroom. The study was carried out by RAYNOISE under the same conditions and acoustical characteristics of field measurements in the classrooms. Several acoustical quantities such as signal-to-noise ratios, reverberation time, and speech transmission index have been studied. Comparison was focused on the qualities of speech communication between the classrooms with and without renovation. Additional comparison was conducted with the results of field measurements. The results of the present study found that the quality of speech communication varied depending on changes in absorption. The results further implied the limitations and the precautions on application of RAYNOISE for different studies. [Work supported by Research Grants Council of The Hong Kong SAR Government, Research Committee of The Hong Kong Polytechnic University, and Architectural Services Department of The Hong Kong SAR Government.]

4:30

3pNSc6. Speech privacy unmasked. David Lubman (14301 Middletown Ln., Westminster, CA 92683) and Louis C. Sutherland (Rancho Palos Verdes, CA 90275)

Background masking noise is widely used to achieve speech privacy between work stations in open offices and in nearby closed (executive) offices. Although speech privacy is often essential, high masking noise levels reduce the abilities of some workers to perform speaking and listening tasks effectively and without strain. Masking noise levels are usually set as high as office inhabitants will tolerate, typically about 45–48 dB (A-weighted). This is 10 to 20 dB higher than noise levels in typical home settings. Many who can communicate effectively in home settings are impaired in offices with background masking noise. People adversely impacted include those with mild or worse hearing disabilities of hearing, speech, and language. Many are older workers and non-native speakers. The current architectural trend to greater inclusiveness underscores the desirability of lower background masking noise levels. Designers can usually achieve speech privacy with lower masking noise levels by specifying partitions with higher noise isolation ratings. Higher sound isolation ratings cost more than higher masking noise levels. But higher costs may be balanced by the benefits of a more inclusive and humane workspace.

4:45

3pNSc7. Objective and subjective evaluation of the acoustic comfort in public classrooms. Paulo H. Zannin, Fabiano B. Diniz, and Daniele Krauser (Mech. Eng. Dept., Federal Univ. of Parana, Rua Benedicto Berillo Fangueiro, 108, Santa Barbara, Curitiba, Parana, CEP, 81540-420, Brazil)

The acoustic comfort of classrooms in a Brazilian public school has been evaluated through interviews with 62 teachers and 464 pupils and measurements of background noise, reverberation time, and sound insulation. Acoustic measurements have revealed the poor acoustic quality of the classrooms. Results have shown that teachers and pupils consider the noise generated and the voice of the teacher in neighboring classrooms as the main sources of annoyance inside the classroom. Acoustic simulations resulted in the suggestion of placement of perforated plywood on the ceiling, for reduction in reverberation time and increase in the acoustic comfort of the classrooms. [Work supported by DAAD, CNPq and FINEP.]

5:00

3pNSc8. Experimental determination of end reflection loss (ERL) over a range of common types and configurations of duct terminations. Alexander P. Michaud and Kenneth A. Cunefare (771 Ferst Dr., Rm. 205, Atlanta, GA 30332)

The two-microphone impedance tube method ASTM E1050 was applied to directly measure the termination impedance of rectangular duct configurations of different aspect ratios and end conditions over a 25- to 500-Hz frequency band. Measured values of the acoustic impedance, and hence the reflection coefficient, of the duct termination permits computation of end reflection loss. ERL is essentially the insertion loss of duct termination. The measurement program is challenging, as the precision required for accurate ERL computation is much higher than that for accurate determination of the reflection coefficient. The measurement program is part of a project to assess the available published data on duct ERL. The ASHRAE Handbook contains a tabulation of ERL data derived from predictions of the radiation impedance of a circular plane piston, and there have been some validation measurements made on a limited set of circular duct configurations and orifice plate terminations. Results to date closely conform to published ERL values. [Work supported by ASHRAE.]

Contributed Poster Papers

Posters will be on display from 1:00 p.m. to 5:30 p.m. Authors will be at their posters from 5:15 p.m. to 5:30 p.m.

3pNSc9. Classroom acoustics: Future criteria for the design and renovation of learning spaces. Sooch San Souci, Line Guerra, Nicolas Teichner (AiA—Audition, Intelligibility, Acoust., 47 rue Le Corbusier, 92100, Boulogne, France, sooch.sansouci@aiacoustique.com), and Dick Campbell (Bang-Campbell Assoc., Falmouth, MA)

Recent literature suggests that some of today's criteria for classroom acoustics might be in need of revision. This article explores the traditional acoustic classroom criteria as they are applied today, and why several of the design goals for construction or renovation of learning spaces might not be appropriate acoustic targets. The acknowledged quest for optimized acoustics for the enhancement of learning is discussed and related to a series of newer metrics for the measure and modeling of speech level, occupied and unoccupied noise, controlled reverberation, and SNR values presented on a seat-by-seat basis. Based on the measured results of classrooms designed and built in accordance to these newer metrics, our research indicates that these guidelines would incur an average increase of

2% to 8% to current budgets. Specifics involving building costs versus acoustic value are also discussed. Further, research involved in the evaluation of scholastic advantages is underway.

3pNSc10. Speech and noise levels in classrooms—signal, noise, reverberation levels as a metric for acoustic design for learning. Line Guerra, Nicolas Teichner, Sooch San Souci (AiA—Audition, Intelligibility, Acoust., 47 rue Le Corbusier, 92100, Boulogne, France, line.guerra@aiacoustique.com), and Dick Campbell (Bang-Campbell Assoc., Falmouth, MA)

This study attempts to investigate the variance of speech levels in occupied classrooms during learning activities and their relationship to the total noise component (occupied, ambient, and reverberant). The methodology for the construction of a valid data base is presented, which would

apply to criteria such as SNR (signal-to-noise ratio) and correlate to the term SNRL (signal, noise, reverberation levels). The interrelationship of speech and noise in any learning environment requires that the common variables associated with a stated speech level value should be stated along with the speech level value. The significance of the influences on the global value and in some cases the spectra of speech is discussed relative to variables such as the activity, the total occupancy and comportment of the audience, the volume of the room, total occupied noise in proximity to the speaker, speech directivity, effects of head diffraction among seated audience, and the distance from the source. Results from measurements taken in this study in actual classrooms are presented with an emphasis on specificity. This research has produced a strategy for the mapping of the speech-noise environment that significantly supports the use of a modeled SNR value as a goal for design.

THURSDAY AFTERNOON, 30 NOVEMBER 2006　　　　　　　WAIANAE ROOM, 1:00 TO 4:15 P.M.

Session 3pPA

Physical Acoustics: Cavitation II

Claus-Dieter Ohl, Cochair

Univ. of Twente, Faculty of Science and Technology, P.O. Box 217, 7500AE Enschede, The Netherlands

Shigeo Hayashi, Cochair

Univ. of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan

Invited Papers

1:00

3pPA1. Shock-wave−inertial-bubble interaction near an elastic boundary. Georgy Sankin and Pei Zhong (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Box 90300, Durham, NC 27708)

The interaction of laser-generated single cavitation bubbles (maximum diameter $D_{max}=1.34$ mm) near a silicon rubber membrane with a shock wave (39–75 MPa in peak pressure and 1.7 μs in compressive pulse duration) is investigated. The interaction leads to directional, forced collapse of the bubble with microjet formation along shock wave propagation direction. Maximum jet penetration into the membrane is produced during the bubble collapse phase with optimal shock wave arrival time and stand-off distance s, a condition that ensures that the forced collapse time of the bubble matches with the compressive duration of the shock wave to generate presumably the maximum axial jet velocity toward the boundary. A maximum penetration depth of 0.5 mm is observed at dimensionless proximity γ $(=2s/D_{max})=0.7$ when the bubble establishes contact with the boundary during expansion. At larger γ (≥ 1.2), jet impact is significantly attenuated by the interposing water layer separating the collapsing bubble and the membrane surface. Such interaction may provide a novel acoustic means for *in vivo* microinjection, applicable to targeted delivery of macromolecules and gene vectors to biological tissues.

1:20

3pPA2. In-liquid plasma processing using high-frequency or microwave irradiation. Shinfuku Nomura and Hiromichi Toyota (Dept. Mech., Ehime Univ, Matsuyama, 790-8577, Japan, nomu@eng.ehime-u.ac.jp)

In-liquid plasma is generated by applying high-frequency (27 MHz) or microwave (2.45 GHz) radiation to hydrocarbon liquids or water. In-liquid plasma is expected to replace gas-phase plasma because the reaction rates for chemical deposition of in-liquid plasma are much higher than those for conventional vapor deposition. Moreover, the cooling effect of the liquid mitigates thermal damage to the substrate. In-plasma generation requires the existence of bubbles because this plasma occurs in bubbles in liquid. Two methods exist for generating bubbles. One is to generate vapor bubbles by heating the liquid. The other is to generate acoustic cavitaion bubbles. Sonoplasma is created when acoustic cavitation is used to generate in liquid plasma. This plasma is not arc-discharged plasma, but rather high-frequency or microwave plasma. Therefore, it is useful in many industrial applications. Using this technology as a chemical reactor, synthesis of diamondlike carbon, polycrystalline silicon carbide, and carbon nanotubes, decomposition of waste oils, and the production of hydrogen can be achieved easily in hydrocarbon liquids. Temperatures were measured from the ratio of emission intensities of H_α and H_β. Plasma temperatures in hydrocarbon liquids are higher than those in water.

1:40

3pPA3. Acoustical and optical characterization of air entrapment in piezo-driven inkjet printheads. Detlef Lohse, Jos de Jong, Roger Jeurissen, Michel Versluis (Phys. of Fluids, Univ. of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands), Hans Reinten, Marc van den Berg, and Herman Wijshoff (Océ Technologies B.V, 5900 MA Venlo, The Netherlands)

Stability of inkjet printers is a major requirement for high-quality printing. However, in piezo-driven inkjet printheads air entrapment can lead to malfunctioning of the jet formation. A voltage pulse applied to a piezoelectric element causes an ink-filled channel to deform, thereby creating a pressure waveform in the ink. Fluid acoustics are involved to guide the waveform energy towards the nozzle and to create pressure and velocity profiles needed for the droplet jetting process. Droplets are jetted every 50 μs, and the typical nozzle size is 30 μm. The piezo actuator is employed to actively monitor the channel acoustics and to identify distortions at

3p THU. PM

an early stage. Modifications of the response of the piezo actuator indicate entrapped air bubbles and these allow us to investigate them. When the signal is employed as a trigger for high-speed imaging, the consequences of the entrained bubbles on the droplet formation can be visualized. [This study has been financed by the Fundamenteel Onderzoek der Materie (FOM) of The Netherlands under Grant. No. 02MFS39 and by Océ Technologies B.V.]

Contributed Papers

2:00

3pPA4. Fully nonlinear model for spherical bubble oscillations between rigid parallel plates. Jianying Cui and Preston S. Wilson (Mech. Eng. Dept., Univ. of Texas, Austin, TX 78712-0292)

Many medical ultrasound applications involve the interaction of intense acoustic waves with bubbles confined by blood vessels or tissue. As a step toward understanding such systems, a bubble confined between two parallel rigid plates was modeled using the image method. The model equation is of Rayleigh-Plesset form, including compressibility of the liquid and accounting for boundary conditions imposed by the plates [Cui *et al.*, Proc. 17th ISNA, edited by Atchley *et al.* (AIP, New York, 2006), pp. 229–232]. An analytical solution for the second harmonic, and fully nonlinear solutions in the quadratic approximation, were reported previously. Here, a hybrid time-frequency algorithm is presented for solving the complete nonlinear model equation. In particular, the effect of the plates is most conveniently dealt with in the frequency domain, whereas the nonlinearity due to the gas law is best handled in the time domain. The procedure was verified for moderate drive amplitudes by comparison with a spectral algorithm for solving the model equation in the quadratic approximation [Cui *et al.*, J. Acoust. Soc. Am. **118**, 1942 (2005)]. Results obtained from the hybrid algorithm will be presented that illustrate the effect of plate separation for different amplitudes and frequencies of acoustic excitation.

2:15

3pPA5. Controlled cavitation in lab-on-a-chip systems. Claus-Dieter Ohl, Ed Zwaan (Phys. of Fluids, Univ. of Twente, Postbus 217, 7500 AE Enschede, The Netherlands), Sèverine Le Gac (Univ. of Twente, 7500 AE Enschede, The Netherlands), and Kinko Tsuji (Shimadzu Europa GmbH, Duisburg, Germany)

Typically cavitation bubble dynamics is studied either in a free, liquid or close to rigid, free or flexible boundaries. Here, we report on a new experimental approach to study cavitation bubbles in confined geometries with sizes comparable to the cavitation bubble (20–100 μm). These fluid environments are fabricated in a polymeric material (PDMS, polydimethylsiloxane) using a conventional molding process from a silicon master (lab-on-a-chip systems). The bubbles are generated by focusing pulsed laser light into the microfluidic systems filled with a light-absorbing dye. The bubble dynamics is recorded stroboscopically and with high-speed photography. We report on bubble dynamics in various geometries such as in triangular, square, and disk-shaped structures. Additionally, we study a cavitation bubble-based micropump, the effect of the channel width on jetting dynamics, and the interaction of the cavitation bubble on suspension cells. The fluid flow is easily visualized by adding microparticles into the liquid. The authors propose that this new experimental approach opens an interesting toolbox in cavitation research.

2:30

3pPA6. Model for the interaction between a bubble and an elastic particle. Todd A. Hay and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029, haymaker@mail.utexas.edu)

At the Providence meeting, a model describing the interaction of a spherical bubble with a rigid spherical particle was presented [Hay and Hamilton, J. Acoust. Soc. Am. **119**, 3408 (2006)]. Here, elasticity of the particle is considered. The model accounts for pulsation and translation and is accurate to fifth order in terms of the ratio of characteristic radius to separation distance. The dynamical equations are derived using the Lagrangian formalism of Ilinskii *et al.* [Proc. 17th ISNA, edited by Atchley *et al.* (AIP, New York, 2006), pp. 303–310]. Bubbles and elastic solid particles are considered with radii on the order of microns and driven at

megahertz frequencies. The radial particle motion is quasistatic under these conditions. For moderate bubble pulsations at which the liquid may be considered incompressible, translational motion of the bubble and particle is insensitive to whether the particle is elastic or rigid. Consideration is also given to truncation of the model at fifth order. Derivation of the model equations to arbitrary order was facilitated by the development of a computer algebra program. The importance of including higher-order terms for small separation distances is illustrated. [T.A.H. was supported by an NSF Graduate Research Fellowship.]

2:45

3pPA7. Deformation of a single bubble in organic solutions by ultrasound. Hiroshi Itami, Shinobu Mukasa, shinfuku Nomura, Hiromichi Toyota, and Yoshiyuki Takahashi (Dept. of Mech. Eng., Ehime Univ., Matsuyama, 790-8577 Japan)

Behavior of a single bubble under an ultrasound resonant state in a hydrocarbon liquid such as n-dodecane or benzene was observed using a stroboscope and a charge-coupled device (CCD) camera. Although sonoluminescence was not confirmed because of bubble stability limitations, bubble deformation was observed at a situation of stability limitation. Deformation occurred when the bubble contracted to its minimum size. Numerical analyses of deformation growth by Rayleigh-Taylor instability of an oscillated bubble were carried out at the surface. The distortion amplitude of all degrees continued oscillating as the bubble expanded or contracted, and the oscillation frequency increased with the degree. The distortion amplitudes are amplified through the afterbounce by the resonance effect if the oscillation frequency of the distortion amplitude of some degree is approximately equal to the frequency of the afterbounce. The distortion amplitude of all degrees, except for amplified degrees, was dampened by the viscosity effect, and if the distortion amplitudes of all degrees were dampened, the bubble was considered to form a sphere and become stable. A phase diagram of stability of an oscillated bubble showed a function of an ambient bubble radius and the amplitude of forcing pressure.

3:00–3:15 Break

3:15

3pPA8. Observe multibubble cavitations in the ultrasonic cleaning tank. Nakagawa Noritoshi, Wu Chaoqun, and Sekiguchi Yasuhisa (Grad. School of Eng., Hiroshima Univ., 1-4-1 Kagamiyama, Higashi-hiroshima, Japan)

A cylindrical ultrasonic cleaning tank radiated by a PZT with a large horn operated at 27.3 kHz is employed to investigate the manifestations of cavitation in the present research. The characteristics of this cleaning system are analyzed using experimental methods. In this cleaning tank, the nonuniformity of vibration velocity on the radiation surface induces the Eckart streaming, which can eliminate the effect of a standing wave on cleaning efficiency. Fresh tap water and degassed water are employed as test liquids in this research. In tap water, conical streaming is observed as the input voltage of the transducer is increased over 100 V. The structure of conical streaming is anatomized. Its shape varies with the change of input voltage. In degassed water, ultrasonic fireworks are observed and special attention is paid to their subsequences, snowflake bubble clouds and water jet.

3:30

3pPA9. Characterization of in-liquid plasma in a continuous liquid flow. Hiromichi Toyota, Shinfuku Nomura, Shinobu Mukasa, Hiroshi Yamashita (Dept. of Mech. Eng., Ehime Univ., Matsuyama, 790-8577 Japan, toyota@eng.ehime-u.ac.jp), Ayato Kawashima, and Tsunehiro Maehara (Ehime Univ., Matsuyama, 790-8577 Japan)

In-liquid plasma is a useful invention because its chemical reaction rate is several thousand times higher than those of conventional plasma techniques. However, in-liquid plasma is not so familiar now because its fundamental characteristics remain unclear. We investigated characteristics of in-liquid plasma in a continuous liquid flow to facilitate its application as a chemical decomposition or synthesis device. N-dodecane is supplied by the pump and a stable liquid flow is set in the rectangular vessel. Because the flow velocity is insufficient, a nozzle is set 2 mm distant from the microwave electrode tip. The average flow velocity of n-dodecane at the nozzle top is 0–19 m/s and the Reynolds number is as high as 1.2×10^4. A 2.45-GHz microwave is introduced by a monopole antenna electrode inserted into a rectangular cavity resonator. In-liquid plasma is generated and observed in the flowing cavitation bubbles generated from the ultrasonic horn tip. The vessel's inner pressure is adjusted to 65–450 hPa. The temperature in the plasma reached 3600 ± 300 K; it was almost independent of the flow velocity because the physical properties of the plasma are influenced only slightly by the acoustic cavitation stream.

3:45

3pPA10. Microbubble shape oscillations excited through an ultrasound-driven parametric instability. Michel Versluis (Phys. of Fluids, Univ. of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands), Peggy Palanchon, Dave Goertz, Nico de Jong (Erasmus MC, Rotterdam, 3000 DR Rotterdam, The Netherlands), Ivo Heitman, Sander van der Meer, Benjamin Dollet, and Detlef Lohse (Univ. of Twente, 7500 AE Enschede, The Netherlands)

An ultrasonically driven air bubble can become shape-unstable through a parametric instability. Here, we report time-resolved optical observations of shape oscillations (mode $n = 2$ to 6) of micron-sized single air bubbles for a range of acoustic pressures. The observed mode number n was found to be linearly related to the resting radius of the bubble. Above the critical driving pressure threshold for shape oscillations, which as expected is minimum at the resonance of the volumetric radial mode, the observed mode number n is independent of the forcing pressure amplitude. The microbubble shape oscillations were also analyzed numerically by introducing a small, nonspherical linear perturbation into a Rayleigh-Plesset-type equation model which includes a physical thermal damping mechanism describing heat and mass transport in the thin boundary layer at the bubble-to-water interface. Indeed, a parametric instability is responsible for the shape oscillations, and the Rayleigh-Plesset-type equation captures the experimental observations in great detail.

4:00

3pPA11. Numerical investigation of interactions between ultrasound propagation and cavitation bubbles in shock wave therapy. Marko Liebler, Thomas Dreyer, and Rainer E. Riedlinger (Institut fuer Hoechstfrequenztechnik und Elektronik, Universitaet Karlsruhe, Kaiserstr. 12, D-76131 Karlsruhe, marko.liebler@ihe.uka.de)

Acoustic cavitation in shock wave applications is responsible for several therapeutic effects as well as for undesired side effects. Using a two-phase continuum approach for bubbly liquids, a numerical model is presented to simulate the interactions between nonlinear ultrasound propagation and cavitation bubbles. Numerical results are presented, showing in what manner the presence of cavitation bubbles alters the pressure pulse propagation and how the bubble dynamics can be specifically affected by changing the shock pressure waveform. An increased bubble density leads to significant changes in the tensile part of the pressure waveform. This effect is caused by the changes in medium compressibility due to the oscillating cavitation bubbles. Calculations for the bubble cloud dynamics induced by three pressure waveforms with identical energy but different temporal forms of the tensile part are compared. Simulation results demonstrate that tensile parts with low amplitude but long duration increase the cavitation effect compared to high amplitude but short tensile phases.

THURSDAY AFTERNOON, 30 NOVEMBER 2006　　　　　　WAIALUA ROOM, 1:00 TO 3:45 P.M.

Session 3pPP

Psychological and Physiological Acoustics: Auditory Grammar

Valter Ciocca, Cochair
Univ. of Hong Kong, 5F Prince Philip Dental Hospital, 34 Hospital Rd., Hong Kong

Kazuo Ueda, Cochair
Kyushu Univ., Dept. of Applied Information and Communication Sciences, 4-9-1 Shiobaru, Minamiku, Fukuoka 815-8540, Japan

Chair's Introduction—1:00

Invited Papers

1:05

3pPP1. Auditory grammar: The event construction model and spoken language. Yoshitaka Nakajima (Dept. of Acoust. Design, Kyushu Univ., 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan, nakajima@design.kyushu-u.ac.jp)

Several auditory illusions are accounted for by using the event construction model as presented by Nakajima *et al.* (2000) in order to explain the gap transfer illusion. In this model, the perceptual system detects and processes onsets and offsets of sounds as if they were independent elements. Temporal edges of sounds and their perceptual attributes are combined to construct auditory events

obeying the proximity principle in time and frequency. For example, the onset and offset of a Chinese syllable belonging physically to different tones can be combined if close enough in time and frequency. The listener often hears a syllable with a lexical tone resulting from this illusory connection. Illusory attribution of perceptual features also occurs in nonspeech tones. One demonstration where only an offset of a glide is presented shows the mechanism; a very short illusory tone with no physical counterpart is perceived. Probably, the perceptual system looks for an onset to connect with the offset, but fails and is forced to restore an onset. Our attempts to connect this model to speech perception and pitch perception are presented along with some novel demonstrations. [Work supported by JSPS and 21st Century COE, Kyushu University.]

<center>1:25</center>

3pPP2. The 100- to 150-ms window of auditory temporal processing and its consequences. Pierre Divenyi (Speech and Hearing Res., VA Medical Ctr., Martinez, CA 94553)

Robert Efron (Percept. Psychophys. **14**, 518–530 [1973]) demonstrated that the minimum duration of a percept is between 100 and 130 ms for auditory, vibrotactile, and visual stimuli, suggesting the presence of a most likely centrally located temporal integrator with a time window of about that length. Single-unit studies in the cat auditory cortex also unveiled an integration time of 100-150 ms (C. E. Schreiner and R. V. Urbas, Hear. Res. **21**, 227–241 [1986]). This window corresponds to the lower limit of the longest among Ira Hirsh's three-tiered temporal processing ranges [in *Sensation and Perception*, edited by H. R. Moskovitz *et al.* (Riedel, Dordrecht, 1974), pp. 251–258]. The present study argues that a temporal window of this duration can account for a variety of auditory perceptual results requiring temporal processing, such as temporal order identification, segregation of dynamically changing streams, robust speech intelligibility, and backward masking. Results also show that this window routinely becomes longer with age, with the possibility that the increased length of the window may account for some well-known central auditory processing deficits in aging. [Work supported by NIH, the VA Medical Research, and the 21st Century COE Progam, Kyushu University.]

<center>1:45</center>

3pPP3. Types of illusory continuity of verbal and nonverbal sounds: Some observations leading to their discoveries. Richard M. Warren (Univ. of Wisconsin—Milwaukee, Dept. of Psych., P.O. Box 413, Milwaukee, WI 53201-0413)

When gaps of a few hundred ms in a signal (e.g., speech, music, steady-state tones, tone glides) are replaced by a louder sound capable of masking the missing portion, the signal appears intact; the absent fragment can be perceptually synthesized through reallocation of a portion of the louder sound's auditory representation. General rules and the special rules governing various types of restoration will be discussed, as well as how they were found through a series of observations, inferences, and experiments. [Work supported by NIDCD Grant No. R01 DC000208, and travel expenses provided by the COE Program of Kyushu University.]

<center>2:05</center>

3pPP4. Music sound restoration and the effect of presentation conditions. Takayuki Sasaki (Miyagi Gakuin Women's Univ., Sendai, Japan)

When a music sound in a melody was deleted and replaced by a noise, a missing sound can be restored perceptually. This illusory phenomenon in audition, named music sound restoration, can be proved indirectly by inaccurate judgment of temporal position of the inserted noise. In the present study, music sound restoration was investigated in terms of the temporal position of replaced noise, headphone listening or free field listening, and the listeners' knowledge of the phenomenon. The results showed that the accuracy of the noise location differed between the positions of replaced noise in the melody or within a beat. The perceptual restoration occurred not only in a headphone listening but also in a free-field listening. It was revealed that the accuracy of noise localization elevated when the listeners knew the stimulus construction, but it was still lower than that of a blank. It was also confirmed that a missing sound could be restored perceptually, not only in a melodic sequence but also in a simple sequence such as scale. [This work was supported by JSPS and 21st Century COE, Kyushu University.]

<center>2:25</center>

3pPP5. An analysis of timing microstructure in Mozart's piano sonata K.331. Kengo Ohgushi (1-6-10 Oheyama, Goryo, Nishikyo-ku, Kyoto 610-1102 Japan, ohgushi-19.38@r7.dion.ne.jp)

It is known that a phrase-final lengthening, including a pause, plays an important role both for the intelligibility and the naturalness of speech. This fact is likely to hold true for performance of music. The present study investigates (a) how the phrase-final lengthening appears in performances of Mozart's piano sonata in A Major (K.331) and (b) what are common and individual features of the performances by five groups of pianists, European professional, Chinese student, Korean student, Japanese professional, and Japanese student pianists. One hundred and twelve pianists' performances of the first eight measures are analyzed in terms of the inter-onset intervals (IOIs). The data are averaged for the respective groups. The IOI patterns indicate (a) the final notes of the first half measures as well as the final notes of the second half measures are lengthened, (b) final note lengthening is more prominent in the fourth measure where the first phrase ends than in the other measures, and (c) Japanese student pianists play the rhythm of dotted eighth note and the sixteenth note differently from the pianists of other countries. [Work supported by JSPS and 21st Century COE, Kyushu University.]

3pPP6. Lost silence as a consequence of stream segmentation. Shuji Mori (Cognit. Sci. Lab., ISEE, Kyushu Univ., 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan)

A sequence of successive tones can be segmented into perceptually distinctive streams by inserting a silent gap or by changing the frequency in mid-sequence. This study demonstrates that a silent gap is perceptually lost when the gap coincides with the frequency change that occurs between tones. Even a 40-ms gap is undetectable when the tonal frequency is shifted from 800 to 3200 Hz. One interpretation of this "lost silence" phenomenon is the prioritization of frequency change as a cue for auditory segmentation. A silent gap and a change in frequency both mark a stream boundary if they are salient. They will be redundant as boundary markers when they coincide with each other; the frequency change will preempt the priority in auditory processing, whereas gap information will receive little processing and eventually become indiscernible. This phenomenon might be related to perceptual ambiguity of the temporal relation of separate streams (Van Noorden, 1975). [Work supported by JSPS.]

Contributed Paper

3:05

3pPP7. Exclusive allocation of a gap that is shared by two separate sounds. Gerard B. Remijn and Yoshitaka Nakajima (Kyushu Univ., Faculty of Design, 4-9-1 Shiobaru, Minami-ku, 815-8540 Fukuoka, Japan)

When a short temporal gap (40 ms or less) is inserted at the crossing point of two separate sounds of unequal duration, the gap is often allocated only to the shorter sound, whereas the longer sound can be perceptually completed (modal completion). This study shows that the exclusive allocation of the gap to the shorter sound—and the perceptual completion of the longer sound—is most profound when the start of the short sound and the start of the gap are separated by 400 ms or less. A possible explanation for the exclusive allocation of the gap is that within this temporal window, attention is engaged in the formation of a new auditory stream, a process initiated by the shorter sound's onset. When subsequent auditory elements—even a gap—arrive within the temporal window of attentional engagement, these elements will be automatically included in the new stream. [This study was supported by the COE program of Kyushu University, Japan, and the Japan Society for the Promotion of Science.]

3:20–3:45
Panel Discussion

THURSDAY AFTERNOON, 30 NOVEMBER 2006   HONOLULU ROOM, 1:30 TO 4:10 P.M.

Session 3pSA

Structural Acoustics and Vibration and Architectural Acoustics: Vibration and Impact Sound in Buildings

Takashi Koga, Cochair
Kajima Corp., 1-2-7 Moto-Akasaka, Minato-ku, Tokyo 107-8388, Japan

James E. Phillips, Cochair
Wilson Ihrig and Associates, Inc., 5776 Broadway, Oakland, CA 94618-1531

Chair's Introduction—1:30

Invited Papers

1:35

3pSA1. Japanese standards and situation of studies about heavy and soft impact sources. Atsuo Hiramitsu (Bldg. Res. Inst., 1 Tachihara, Tsukuba-City, Ibaraki 305-0802, Japan, hiramitu@kenken.go.jp) and Katsuo Inoue (Nihon. Univ, Kanda-Surugadai 1-chome, Chiyoda-ku, Tokyo 101-8308, Japan)

This report describes standards of heavy and soft impact sources. They are used for measurement of floor impact sound insulation of buildings. In Japan, the method for field measurement of floor impact sound level (JIS A 1418) was revised in 2000 and separated into two parts. Part 1 uses a standard light impact source and corresponds to ISO 140-7. Part 2 uses heavy and soft impact sources. At that time, a rubber ball was standardized along with a car-tire source in JIS A 1418-2. Moreover, JIS A 1440-1 and JIS A 1440-2 for laboratory measurement of the reduction of transmitted floor impact sound have been prepared. JIS A 1440-1 corresponds to ISO 140-8, and JIS A 1440-2 uses heavy and soft impact sources. Floor impact sources characteristics are introduced. They include force characteristics, temperature dependence, stability, and others. In addition, among contents of JIS A 1440-2, the measurement method, calculation method, and a laboratory of a box-frame-type reinforced concrete construction are explained. In addition, study examples using a rubber ball are introduced. As examples, we present the·floor impact sound level, the transmitted floor impact sound level, hearing tests on floor impact sound, and evaluation of walking sensations.

1:55

3pSA2. Trends for prediction methods of heavyweight floor impact sound. Takashi Koga (Bldg. Construction Management Div., Kajima Corp., 1-2-7 Moto-Akasaka, Minato-ku Tokyo 107-8388, Japan, tkoga@kajima.com), Katsuo Inoue (Nihon Univ., Tokyo 101-8308, Japan), and Shinji Nakazawa (Tekken Corp., Chiba, 286-0825, Japan)

Trends of building structures of Japanese multifamily dwellings and prediction methods of heavyweight floor impact sound are introduced. Floor impact sound-reduction performance caused by human action like stepping and jumping is considered to be a vital function for Japanese apartment complexes. There are quite a few cases where the floor thickness shall be designed from 200 to 300 mm because of acoustical performance, to assure the required performance. The practical prediction method, the so-called impedance method, was been established in the 1970s, and it has been commonly used. Though there are many buildings that are out of the applicable scope of the ordinal impedance method with the rise of structural design technology, the outline shown is that of the revised practical prediction method that is under discussion by the working group of the Architectural Institute of Japan. And, examples of hybrid method are also introduced that substitute the impedances by numerical vibration calculation results using finite-element method for simplified calculation results of impedance method.

2:15

3pSA3. Correlation of field measurements for footstep force pulse with finite element analysis model for determining vibration response of a building floor. Richard Carman and James Phillips (Wilson, Ihrig & Assoc., 5776 Broadway, Oakland, CA 94618)

An idealized force pulse that simulates footsteps was developed by Galbraith and Barton. Their proposed forcing function was determined from direct force measurements using walkers of different weights walking at various speeds. Unger and White used this idealized footstep pulse to determine the maximum dynamic deflection of a floor for the purpose of building design calculations. These basic concepts were used to revisit the characteristics of the idealized forcing function. During the design phase of an academic research facility, extensive vibration measurements were obtained using an exemplar floor similar in design to that proposed for the new research building. The exemplar floor was modeled using FEA. Its modal characteristics were compared to the field test results and found to be in reasonable agreement. Upon having established confidence in the basic FEA model, the field measurement results of two walkers on the exemplar floor were compared with the FEA results predicted using the Galbraith-Barton footstep pulse. The shape and magnitude of the footstep pulse were iteratively adjusted until the measurement results and FEA results were in good agreement. The results of the FEA analysis and the measurement program behind the model will be presented.

2:35

3pSA4. Influence of lightweight structural design of office buildings on vibration levels due to people walking and operation of mechanical equipment. Steven D. Pettyjohn (The Acoust. & Vib. Group, 5700 Broadway, Sacramento, CA 95820-1852)

West coast buildings typically use lightweight structural designs, even for Class A office building. This means the best floors use lightweight concrete on fluted decks with beam and column spacings of 30 to 35 ft. with open web joists at 10 ft. on center. Class B office building may use columns at 28 and 35 ft. with only two-layers of plywood on wood open web joists at 24 in. on center. The resulting floor/ceiling assemblies are structurally sound but dynamically very live. Some structural damping would be expected from the introduction of partitions, but with the use of large open areas and partial height partitions, damping is limited. These assemblies have relatively low resonance frequencies and are very susceptible to any sinusoidal input force such as people walking or operation of mechanical equipment. Vibration measurements, made on four buildings after complaints were received, show the source of some of these problems. The results are compared with predictions by Murray *et al.* and Ungar *et al.* The lack of composite action between the beams and joists/purlins and the floor appears to be the source of many of the problems. Solutions to the problems in these existing and occupied buildings are discussed.

2:55−3:10 Break

Contributed Papers

3:10

3pSA5. Heavy and light impact sources to rate impact sound transmission and changes due to applied floor toppings. Alf Warnock and Trevor Nightingale (Acoust. Group, Inst. for Res. in Construction, Natl. Res. Council Canada, Ottawa, K1A 0R6, Canada)

The magnitude and spectrum of the power injected by an impact source depends on the impedance of the source and floor. If standardized impact tests are meant to give ratings that correlate well with subjective impressions of footstep noise, it follows that the impact source used should have the same impedance as an average human foot at least over the range of test frequencies. The ISO tapping machine, the Japanese tire machine, and an 18-cm-diam rubber ball do not satisfy this criterion. Consequently, their impact spectra differ from those from a live walker. Floor toppings, in particular, are ranked differently. Examples of discrepancies will be presented for direct transmission between vertically separated rooms. For horizontally and diagonally separated rooms, flanking transmission controls the impact sound pressure level. The sound pressure level depends not only on the power injected by the source, but also source location relative to the flanking junction because of propagation attenua-

tion across the floor. Sensitivity to source location is similar for different sources (ISO tapping machine and Japanese ball), suggesting that the same source could be used for direct and flanking transmission measurements if the source adequately simulates the impedance of a human foot.

3:25

3pSA6. Evaluation of mid- and high-frequency impact insulation methods. John J. LoVerde and David Wayland Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404)

Impact insulation is currently measured using impact insulation class (IIC) and a standard tapping machine described in ASTM E492 and E1007 (and corresponding ISO metrics and standards). The resultant IIC value does not correlate well with subjective reaction. In previous papers, the authors have focused on the low-frequency thudding from footfalls in lightweight wood construction and shown that the impact sound pressure level (ISPL) in the 63-Hz octave band correlates better with subjective reaction than IIC or other metrics [LoVerde and Dong, J. Acoust. Soc. Am. **112**, 2201 (2002); INCE Inter-Noise 2004 Proceedings, 167 (2004); J. Acoust. Soc. Am. **119**, 3220 (2006)]. Further, the low-frequency ISPL is

controlled by different variables than the mid/high-frequency ISPL. This suggests that a two-metric system would be valuable where the first metric describes the low-frequency performance and the second metric describes the mid/high-frequency performance. Field testing experience is presented showing that IIC does not adequately differentiate the mid/high-frequency performance of various resilient materials. Candidate mid/high-frequency metrics are evaluated to determine if any result in improved methods for evaluation and rank-ordering of floor/ceiling assemblies.

3:40

3pSA7. Flanking impact sound transmission in wood-framed multifamily construction. David Quirt and Trevor Nightingale (Acoust. Group, Inst. for Res. in Construction, Natl. Res. Council Canada, Ottawa, K1A 0R6 Canada)

This paper reports findings from a recently completed study of flanking sound transmission involving the wall/floor junction in wood-framed multifamily buildings. Flanking transmission exists in all buildings, but it can be controlled through good design. The paper examines how common construction details affect flanking paths between horizontally, vertically, and diagonally separated rooms. Variables considered include mounting conditions and orientation of the floor joists, framing of the wall assembly and any associated fire blocking, and mounting and number of layers of gypsum board. Estimates of the apparent sound insulation were obtained by summing the energy transmitted by the direct path through the wall or floor assembly separating the pair of rooms, and all the flanking paths involving the relevant wall-floor junctions. Results indicate that if there is no floor topping (i.e., the subfloor is bare) the apparent sound insulation

for both airborne and impact sources is typically limited by flanking transmission involving the floor. Since, the source room floor is an element common to all impact flanking paths, three different toppings were evaluated as treatment options, and additional layers of gypsum board and resilient mounting were considered as options for the walls. The effectiveness of each option is discussed

3:55

3pSA8. Case history building structural vibration isolation for 500 Atlantic Avenue. Gregory Tocci (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776), George Wilson, James Phillips (Wilson Ihrig & Assoc., Inc.), Gladys Unger, and James Moore (R.H. Lyon Div. of Acentech)

Vent Building 3 of the new Central Artery Tunnel located at 500 Atlantic Avenue in Boston, Massachusetts is a heavy poured in place concrete structure with its roof situated about 3 ft. below grade. The vent building contains 23 100+-in.-diameter double-width centrifugal fans that, for a number of reasons, are not vibration isolated. The fans exhaust into a cluster of reinforced concrete shafts that rises up to about 250 ft. above grade. The tall stacks have been embedded in a luxury hotel/residential condominium complex that Extell Development Company is currently constructing. The stacks provide both vertical support and seismic support for the building. This paper summarizes the history of work on the project, and compares predicted and measured sound and vibration levels in representative building spaces. [We wish to acknowledge the support of Extell Development Company, Boston, MA for their support of this work.]

THURSDAY AFTERNOON, 30 NOVEMBER 2006 LANAI ROOM, 1:00 TO 5:15 P.M.

Session 3pSC

Speech Communication: Speech Timing and Pronunciation Training for the Japanese Language

Robert F. Port, Cochair
Indiana Univ., Linguistics, 330 Memorial Hall, Bloomington, IN 47405

Yukari Hirata, Cochair
Colgate Univ., Dept. of East Asian Languages and Literature, 13 Oak Dr., Hamilton, NY 13346

Chair's Introduction—1:00

Invited Papers

1:05

3pSC1. Timing variations of frequent utterances occurring in a very large corpus of spontaneous Japanese conversational speech. Nick Campbell (ATR, Kyoto, Japan)

A corpus of spontaneous conversational Japanese speech was collected from volunteer subjects who wore high-quality head-mounted microphones and recorded their daily spoken interactions to minidisk over a period of 5 years. All recordings were transcribed and tagged according to interlocutor type, and a portion representing about 10% was further annotated for speech-act, speaker-state, emotion, and speaking style. This paper presents timing data from the corpus, showing how the same utterance can vary according to speaking style and other factors. It presents the hundred most common utterances in the corpus and relates their durations to spectral and prosodic characteristics that vary according to affect, attitude, intention, and relationship with the listener.

1:25

3pSC2. On the generation and perception characteristics of Japanese timing control. Yoshinori Sagisaka (Waseda Univ. 29-7 Bldg., 1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo 169-0051 Japan, sagisaka@giti.waseda.ac.jp), Makiko Muto (Univ. of Tokyo, Japan), Hiroaki Kato (NiCT/ATR, Japan), and Minoru Tsuzaki (Kyoto City Univ., Japan)

We have been studying temporal characteristics of Japanese for decades. Not only acoustic measurements but also perceptual studies on temporal modification have revealed the control principles lying behind manifestation of segmental duration characteristics. This talk tries to introduce familiar generation principles such as "mora-timing" and phrasal tempo reset from the viewpoint of humans temporal organization through the understanding of duration statistics. Moreover, by showing perceptual characteristics

including sentential factors, we would like to propose how linguistic units relate to temporal perception and what kind of information form can be considered in temporal cognition process. Some of our recent efforts will also be presented to investigate language-dependent temporal characteristics by observing non-native learners temporal control characteristics. We really expect that we will be able to have a better picture of generation and perception mechanism through the cross check of their characteristic differences between natives and non-natives. [Work supported by Waseda Univ. RISE research project of "Analysis and modeling of human mechanism in speech and language processing" and Grant-in-Aid for and Scientific Research A-2, No. 16200016 of JSPS.]

1:45

3pSC3. Temporal control in Japanese dialects. Haruo Kubozono (Kobe Univ., Nada-ku, Kobe 657-8501 Japan, kubozono@lit.kobe-u.ac.jp)

This paper discusses the extent to which temporal control varies among Japanese dialects. Most dialects of Japanese including Tokyo Japanese are so-called "mora dialects" in which the mora serves as the basic unit of timing. In the periphery of Japan, however, we find a couple of dialects called "syllable dialects" whose basic phonological unit is the syllable rather than the mora. Kagoshima Japanese (KJ) is one such dialect, but little is known about its temporal organization. The main purpose of this paper is to report on the results of an acoustic experiment on this syllable-based dialect and to compare them with those of the experiments about Tokyo Japanese (TJ). This comparison reveals the following four points: (*i*) geminate consonants are considerably longer than single consonants in both TJ and KJ, (*ii*) however, the phonetic duration of words in KJ is not determined by the number of moras involved but by the number of syllables, (*iii*) in KJ, vowels are consistently shorter when they follow geminate consonants than when following single consonants, and (*iv*) unlike English, Korean, and other languages, neither TJ nor KJ shows an effect of vowel shortening in closed syllables.

2:05

3pSC4. Japanese mora-timing and processing: The case of devoiced vowels. Natasha Warner and Naomi Ogasawara (Dept. of Linguist., Univ. of Arizona, P.O. Box 210028, Tucson, AZ 85721-0028, nwarner@u.arizona.edu)

An extensive literature has argued that Japanese speakers compensate for inherent durational differences to keep moras to approximately equal lengths (e.g., Han, 1962; Port *et al.*, 1980; Minagawa-Kawai, 1999; reviewed in Warner and Arai, 2001a). However, near-regularity of moras may not be because of intentional compensation (Warner and Arai, 2001b). This paper examines Japanese listeners processing, rather than production, of one case of irregularly timed moras: vowel devoicing. When Japanese vowels are devoiced, the mora tends to be shorter than usual. If speakers aim for regular timing, one might predict that listeners would also expect moras to be regularly timed and would thus have difficulty processing sequences that deviate from regular mora timing. However, one could predict that listeners are aware of what deviations from regular timing to expect, and thus have no difficulty processing timing deviations. The results, from phoneme monitoring and lexical decision tasks, show that listeners only have difficulty processing devoiced vowels in environments where devoicing would not be expected. In fact, where devoicing would be typical, listeners find devoiced vowels easier to process. This suggests that in processing, knowledge of patterns in the language outweighs any bias toward regularly timed moras.

2:25

3pSC5. Perception and production of long and short vowels in Japanese by children. Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan), Kyoko Iitaka, and Eriko Ohki (Sophia Univ., Chiyoda-ku, Tokyo, 102-8554 Japan)

Studies have shown that acquisition of reading depends on such phonological processes as verbal short-term memory and temporal processing of speech. An inability to encode phonological representations orthographically may cause reading disabilities, such as dyslexia. Categorical perception is known to be closely associated with phonological encoding. Thus, studies examining the temporal processing of speech sound categorization may be useful in differential diagnoses of reading development and disabilities. The first study examines the perception of long and short Japanese vowels by a group of seven normal six-year-old children and two with delayed phonological development. Subjects were asked to identify vowel length in sets of words. The first vowel in each stimulus word was made to vary with respect to length. Perception by normal and delayed children differed in ways similar to what had been reported in previous English studies. In the second study, the relation between perception and reading of long/short vowels was examined by a group of 125 children ages 4 to 8. Similar settings were used for the second experiment. The results showed that (1) a large developmental change of perception was observed in ages 6–8 and (2) a positive correlation was obtained between perception and reading tasks.

2:45

3pSC6. Effects of speaking style on the regularity of mora timing in Japanese. Michael Connolly Brady, Robert F. Port, and Kyoko Nagao (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405)

The regularity of mora timing in Japanese has remained controversial over the years [Warner and Arai, Phonetica **58**, 1–25 (2001)]. It is possible that the degree of regularity varies with speaking style. Four Japanese subjects spoke six-mora proper names with varying syllable structures where the second name had either three simple syllables (e.g., *Tomiko*) or two syllables with a long vowel (e.g., *Tooko*) or a long consonant (e.g., *Tokko*). These were read either in formal sentences, read in conversational style sentences, or used in spontaneous description of pictures of characters having these names. Moras were measured as the intervals between vowel onsets using an automatic vowel-onset detection algorithm. Although all styles suggested regular mora timing, the results show that the styles differ in the degree of temporal compensation for the moras constituted by long vowels and long consonants which are shorter than consonant-vowel syllables. The compensation was clearest in the most formal style of speech.

3:05

3pSC7. Representations of within-word structure and Japanese speech timing. Takashi Otake (E-Listening Lab. and Max Planck Inst. for Psycholinguist., 3-35-20, Higashi-Tokorozawa, Tokorozawa, Saitama, 359-0021 Japan)

One complication about the debate of mora timing is that durational units are closely related to syllable structure in which morae can be treated as subunits of syllables. As a consequence, durational properties of the components in three types of syllables, CVN, CVV, and CVQ, can equally be examined from the point of representations of within-word structure. Durational properties in Japanese speech timing undoubtedly play an important role, but knowledge of representations of within-word structure cannot be neglected because it is uniquely language specific. In this presentation it is discussed how knowledge of representations of within-word structure plays a significant role in recognition of morae from the point of psycholinguistic framework, presenting various types of data that were collected from (1) Japanese preschool children (before and after acquiring kana orthography), (2) Japanese monolinguals and Japanese English bilinguals, and (3) Japanese L2 learners (from beginning to advanced whose native language is English). Also reported are data from Telugu (a Dravidian language spoken in India), whose speech segmentation mechanism behaves like that of mora-timed language. It is hoped that knowledge of representations of within-word structure together with the durational properties will facilitate Japanese learners to acquire Japanese speech timing.

3:25–3:55   Break

3:55

3pSC8. DiscourseWare, a blended learning courseware for understanding and producing Japanese audio-visual reactive tokens. Goh Kawai and Takafumi Utashiro (Inst. of Lang. and Culture Studies, Hokkaido Univ., Sapporo, Hokkaido 060-0817 Japan, goh@kawai.com)

We developed blended learning courseware called DiscourseWare for learning Japanese audio-visual reactive tokens. Acquiring proper token choice, timing, frequency, duration, and intonation are emphasized. The courseware consists of interactive computer-based multimedia material and classroom instruction. Lessons start with instructor-led classroom instruction on the types and usage of various reactive tokens. Learners then independently use interactive computer-based courseware to learn how to choose appropriate reactive tokens. The courseware contains videos showing two native speakers conversing using a uncontrolled vocabulary and syntax and at a natural speed. The reactive token being taught is blanked out. Learners solve multiple-choice problems asking which reactive token would be appropriate in that context. Learners increase proficiency in interpreting various prosodic patterns of the same phonetic string. After computer-based training, learners regroup and practice exchanging reactive tokens. Our courseware is accompanied by a syllabus for reactive tokens that might interest researchers of comparative pragmatics as well as language instructors. The conference presentation will include DiscourseWare demonstrations, and before-and-after training comparisons of non-native learners' reactive token productions as assessed through interviews with language teachers.

4:15

3pSC9. An issue of rate normalization for learners of Japanese. Yukari Hirata (Dept. of East Asian Lang. and Lit., Colgate Univ., 13 Oak Dr., Hamilton, NY 13346)

This talk explores native and non-native speakers' identification of Japanese phonemic length contrasts in sentences of varying speaking rates. Native Japanese speakers were found to use the rate of the sentence surrounding a target word as a perceptual cue (i.e., to normalize speech rate) to distinguish phonemic vowel and consonant length. This perceptual cue overrode more local cues such as physical duration of the target vowel or consonant. However, when beginning learners of Japanese participated in the same listening tasks, they were unable to normalize speech rate and identify length of vowels and consonants as native speakers did. Specifically, when asked to identify the length of vowels in /mV(V)mV(V)/ (with the first vowel accented) in sentences of slow and fast rates, the learners were least accurate in identifying the long vowel of /mVmVV/ in sentences spoken at a fast rate. The learners' perception was biased by physical duration of vowels and by the presence or absence of pitch accent. Their identification accuracy for long vowels was more affected by speaking rate than that for short vowels. Taken together, these results suggest a need for a training program to help learners of Japanese to cope with speaking rate variations.

4:35

3pSC10. Effects of exposure and training on perception of Japanese length contrasts by English listeners. Keiichi Tajima (Dept. of Psych., Hosei Univ., 2-17-1 Fujimi, Chiyoda-ku, Tokyo 102-8160, Japan, tajima@i.hosei.ac.jp), Hiroaki Kato (NICT/ATR, Kyoto 619-0288, Japan), Amanda Rothwell, Reiko Akahane-Yamada (ATR, Kyoto 619-0288, Japan), and Kevin G. Munhall (Queen's Univ., Kingston, ON, Canada)

Native English listeners are known to have difficulty distinguishing Japanese words that contrast in phonemic length, often realized as a contrast in vowel or consonant duration. The present study reports results from a series of experiments investigating the extent to which English listeners' perception of such length contrasts can be modified with exposure to Japanese and with perceptual identification training. Listeners were trained in a minimal-pair identification paradigm with feedback. A pretest and posttest were also administered, using natural tokens of Japanese words containing various vowel and consonant length contrasts, produced in isolation and in a carrier sentence, at three speaking rates, and by multiple talkers. Results indicated that exposure and perceptual training substantially improved identification accuracy. Even though listeners were trained to identify words contrasting in vowel length only, performance also improved for other contrast types. Furthermore, speaking rate strongly affected performance, but training improved performance at all three rates, even though listeners were trained using only stimuli spoken at a normal rate. These results suggest that non-native listeners are highly susceptible to factors that affect the temporal characteristics of speech, but their perceptual strategies can be modified with exposure and training. [Work supported by NICT and JSPS.]

3pSC11. Categorization of phonemic length contrasts in Japanese by native and non-native listeners. Hiroaki Kato (NICT/ATR, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan, kato@atr.jp), Keiichi Tajima (Hosei Univ.), Amanda Rothwell, Reiko Akahane-Yamada (ATR), and Kevin Munhall (Queen's Univ.)

Previous studies reported that exposure to Japanese and identification training allowed English listeners to improve their identification accuracy of Japanese words with phonemic length contrasts. However, this does not necessarily mean their perceptual properties approach those of native listeners in all respects. Past perceptual studies have shown that native listeners show, for example, an extremely sharp short-to-long boundary and shift of the boundary to adapt to changes in the temporal context, i.e., typically speaking rate variations. To carefully investigate the effects of training on such properties, the present study analyzed the learners' identification of stimulus continua between word pairs that minimally differed in the length of a phoneme. Overall results showed that English listeners' boundaries tended to be sharpened by the identification training, but only to a limited extent. On the other hand, the results also showed that English listeners' boundaries shifted across different speaking rates in ways that resemble native listeners' adaptation tendencies. The latter result suggests that even non-native listeners can adjust their identification boundaries according to differences in temporal context. Discussion will include both prospects and limitations of listeners' improvement by the perceptual training in this study. [Work supported by NICT and JSPS.]

THURSDAY AFTERNOON, 30 NOVEMBER 2006   MOLOKAI ROOM, 1:00 TO 5:00 P.M.

Session 3pSP

Signal Processing in Acoustics: Various Topics in Signal Processing (Poster Session)

Charles F. Gaumond, Cochair

Naval Research Lab., Code 7142, 4555 Overlook Ave., S.W., Washington, DC 20375-5320

Yoshinobu Kajikawa, Cochair

Kansai Univ., Dept. of Electronics, Faculty of Engineering, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan

Contributed Papers

All posters will be on display from 1:00 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 5:00 p.m.

3pSP1. Adaptive surround technology for stereo music signals. Yoshitaka Murayama, Akira Gotoh, Jun Ota, Hareo Hamada (DiMAGIC, Sugisho Bldg 3F, 2-4-5 Kanda-Ogawamachi, Chiyodaku, Tokyo 101-0052, Japan), Yuko Watanabe, and Hareo Hamada (Tokyo Denki Univ., Inzai-shi, Chiba, Japan)

Computation of the left and right channel instantaneous difference $(xL - xR)$ is mostly used to obtain the signals for surround sound. This signal is usually referred as the antiphase components. However, there are many problems to using the antiphase components in terms of its applicability in practice as well as the sound quality. For instance, when signals xL and xR are equal, or almost equal, this variation, that is to say, the correlation coefficient value, is high; there is no way to determine the direction vector uniquely, therefore an ambiguity may appear. Instead of the antiphase component's computation, a new method is proposed by using an adaptive signal processing technique. In this paper, it is revealed that natural surround sound signals having no antiphase and uncomfortable feeling are adaptively generated by the proposed method. The other advantages from a signal processing point of view are also discussed. Finally, subjective evaluations on sound quality of a surround sound system are carried out.

3pSP2. A design of fast steering filters based on the adaptive fusion of predesigned finite impulse response filters for microphone array. Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, gr020040@se.ritsumei.ac.jp), Takanobu Nishiura, and Yoichi Yamashita (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

With a microphone array, a desired distant-talking speech signal can be acquired selectively by steering the directivity in noisy environments. To form directivity, adaptive beamformers have been proposed as conventional beamformers. They can form null directivity with a small number of transducers. However, adaptive beamformers have a serious drawback: it can't realize the noise-robust adaptation in variable noisy environments, because it has necessary heavy computational costs to train multi-channel adaptive steering filters. To solve this problem, we propose to design the fast steering filters based on the adaptive fusion of predesigned FIR filters in variable noisy environments. The adaptive fusion of predesigned FIR filters is realized by adding (or subtracting) various predesigned FIR filters immediately. Therefore, the proposed method can quickly design the steering filters even in variable noisy environments. In addition, microphone calibration is indispensable for applying the proposed method. Therefore, we also propose the new microphone calibration method based on environmental noise without target signal. Evaluation experiments were con-

ducted in a real acoustic room. As a result of evaluation experiments, we could confirm that the ASR (automatic speech recognition) performance of distant-talking speech was improved on the well-known condition of noise source directions. [Work supported by MEXT of Japan.]

3pSP3. A frequency domain method to estimate the coefficients of feedback control filter. Yusuke Iwamatsu, Masahiro Itou, and Kensaku Fujii (Div. of Comp. Eng., Univ. of Hyogo, 2167 Shosha, Himeji, 671-2280, Japan, er05i005@steng.u-hyogo.ac.jp)

This study proposes a method for repeatedly updating coefficients of a feedback control filter used for canceling a feedback path from a loudspeaker to a noise detection microphone. The coefficients are typically estimated by feeding a sequence of extra noise to the loudspeaker. However, feeding such extra noise under active noise control is undesirable. The proposed method estimates the coefficients without feeding the extra noise. Using the estimation errors involved in the coefficients of the noise control filter instead of the extra noise, the proposed method provides two independent equations and estimates coefficients of the feedback control filter by solving those. The two independent equations are provided through linear prediction using an auxiliary filter. A solution of the equations with two unknowns (a primary noise and the feedback path) provides the coefficients of the feedback control filter, perfectly canceling the feedback path. A salient drawback is that the feedback component is extremely small in comparison to the primary noise that is incident on the noise detection microphone. For that reason, estimation of the coefficients is difficult. This method solves the problem by first estimating the primary noise and then the feedback path.

3pSP4. A new multichannel adaptive algorithm. Masahiro Itoh, Kensaku Fujii (Div. of Comput. Eng., Univ. of Hyogo, 2167 Shosha, Himeji, 671-2280 Japan, er05a003@steng.u-hyogo.ac.jp), and Mitsuji Muneyasu (Kansai Univ., 3-3-35 Yamate-chou, 564-8680 Suita, Japan)

This study presents a new multichannel adaptive algorithm that identifies unknown systems without superposing independent components on reference signals. A problem in such multichannel systems is that the reference signals involve strong cross-correlative components between themselves. These components delay the convergence of the coefficients of adaptive filters. At worst, they can render identification impossible. Usually, this problem is solved by superposing slightly independent components on the reference signals. Various superposition methods are therefore proposed. However, this superposition is perceivable as discordant noise. This study hence examines a new algorithm that exploits the residual components obtained by applying a linear prediction technique to the reference signals. The cross-correlative components thereby decrease and the convergence is hastened. Using this algorithm, however, the residual components are not fed into the unknown systems and adaptive filters. This is a point of great difference from conventional methods. This study also analyzes the performance of the proposed algorithm and derives the convergence condition as a range of step size. Finally, computer simulations are used to verify the derived convergence condition and the increased convergence speed.

3pSP5. Frequency domain simultaneous equations method for active noise control systems. Tadashi Ujino, Kensaku Fujii (Div. of Comput. Eng., Univ. of Hyogo, 2167 Shosha, Himeji, 671-2280 Japan, er06e006@steng.u-hyogo.ac.jp), Mitsuji Muneyasu (Kansai Univ., 564-8680, Suita, Japan), and Yusuke Fujita (Catsystem Corp., 671-1121, Himeji, Japan)

This study proposes a new method for feedforward-type active noise control systems. This method, named the frequency domain simultaneous equations method, is based on a different principle from the filtered-x algorithm requiring a filter modeled on a secondary path from a loudspeaker to an error microphone. Instead of the filter, this method uses an auxiliary filter identifying the overall path consisting of a primary path, a noise control filter, and the secondary path. As seen from the configuration of the overall path, the auxiliary filter can provide two independent equations when two different coefficient vectors are given to the noise control filter. In practical use, the two different coefficient vectors are provided by their own estimation errors. By solving the independent equations, the proposed method can estimate the coefficient vector of the noise control filter minimizing the output of the error microphone. This study first presents computer simulation results demonstrating that the proposed method can automatically recover the noise reduction that is in effect degraded by path changes. Finally, this study applies the proposed method to an experimental system and, by using the system, verifies that the proposed method is functional in practical systems.

3pSP6. Partial differential equation (PDE)-based method of sound source localization. Shigeru Ando and Nobutaka Ono (Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan, ando@alab.t.u-tokyo.ac.jp)

We describe the problem of direction and distance estimate of sound sources in free 3-D space. We first derive a partial differential equation (PDE), $\nabla f(\mathbf{r},t) = \{(1/R)f(\mathbf{r},t) + (1/c)\dot{f}(\mathbf{r},t)\}\mathbf{n}(R \equiv |\mathbf{r}|, \mathbf{n} \equiv -\mathbf{r}/R)$, what we call the sound source constraint (SSC). We show that the general solution of the SSC is a diverging spherical wave from a point sounce with arbitrary temporal waveform. The SSC enables the observer to determine the source location (distance R and direction $\mathbf{n}$) from local measurements of the wave field. As the measurements of the wave field, we consider two methods: (1) the weighted temporal integral of arrayed microphone outputs in a finite duration and (2) the weighted spatial integral of pressure distribution in a rectangular or circular area. In both cases, we obtain exact formulas for localizing a single source from a single measurement and multiple sources from the combination of differently weighted measurements.

3pSP7. Effects of microphone and loudspeaker directivity on accuracy of synthesized wave fronts in sound field reproduction based on wave field synthesis. Toshiyuki Kimura, Sadanori Ito, Masaki Nakagawa (Dept. of Comp. and Info. Sci., Tokyo Univ. of Agriculture and Technol., 2-24-16 Naka-cho, Koganei-shi, Tokyo, 184-8588 Japan), and Kazuhiko Kakehi (Chukyo Univ., Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan)

Sound field reproduction based on wave field synthesis proposed by Camras [M. Camras, J. Acoust. Soc. Am. **43**, 1425–1431 (1968)] is a technique that synthesizes wave fronts of a control area at a listening area using directional microphones and loudspeakers placed at the boundary of those areas. This study evaluates the effects of the directivity of microphones and loudspeakers on the accuracy of synthesized wave fronts. Microphones and loudspeakers were placed at the boundary of both areas that have two cases (a circle with a radius of 2 m and a square 4 m on a side). The number of microphones and loudspeakers were 630 (circle) and 801 (square). The interval of microphones and loudspeakers was less than 2 cm. Therefore, spatial aliasing does not occur if the frequency is less than 8 kHz. Three directional patterns (omnidirectional, unidirectional, and shotgun) were designed as directivity conditions of microphones and loudspeakers. Computer simulation results showed that there is no effect of directivity of loudspeakers and that accurate wave fronts can be reproduced when a unidirectional or shotgun microphone is applied. [Work supported by special grant for education and research of Tokyo Univ. of Agriculture and Technol.]

3pSP8. A study of pinna effect on head-related transfer functions.
Hiromi Sekido, Yuko Watanabe, and Hareo Hamada (Tokyo Denki Univ., 2-1200 Muzai-gakuendai, Inzai, Chiba Prefecture 270-1382, Japan)

For the sound localization, an interaural difference is important. To obtain the desired localization, it is necessary to use a head-related transfer function (refer to HRTF), which is defined as the transfer function between a sound source and a surface of the ear drum. In our previous work, as far as in the case of reproducing the virtual sound source in the horizontal plane, it is confirmed to be capable of remarkable localization by using the numeric data calculated by a rigid sphere head model instead of a dummy head microphone. However, it is necessary to solve the problem of the front back confusion and the sound localization of the vertical plane. For that, it might be essential to add the pinna information to the rigid sphere model, to examine how each part of the pinna is related to the localization.

3pSP9. Sound reproduction system with switching control band range on every listening point.
Yosuke Tatekura and Takeshi Watanabe (Dept. of Eng., Shizuoka Univ., 3-5-1 Johoku, Hamamatsu, Shizuoka, 432-8561, Japan)

We propose a robust multichannel sound reproduction system that utilizes the relationship between the width of the actual listening area and the control frequency of the listening points. The conventional multichannel sound reproduction system uses inverse filters that are designed based on room transfer functions (RTFs) from loudspeakers to listening points in the listening room. However, the sound reproduction accuracy is deteriorated by room environment variations when fixed inverse filter coefficients are used because the condition number relevant to the instability of the matrix composed of the RTFs is large, especially at low frequencies; the inverse filters are unstable at those frequencies. This tendency is more remarkable when listening points are arranged more closely. To resolve the problem, we switch the control band at every listening point to avoid degrading the reproduced sound in low-band frequencies. The control band range of the listening points at both ears is only high band. That of the other listening points is the entire control band. Numerical simulation with real environmental data showed that improvement of the reproduction accuracy is about 6.4 dB on average, even with temperature fluctuations of 5 °C as environmental variations of the listening room.

3pSP10. Frequency scaling reduces individual differences in external-ear transfer functions in developing cats.
Daniel J. Tollin (Dept. Physiol & Biophys., Univ. Colorado Med. Sci., 12800 E. 19th Ave., Aurora, CO 80045, daniel.tollin@uchsc.edu)

There are three cues to sound location: Interaural differences in time and level and monaural spectral shape cues. The ultimate magnitudes of these cues depend largely on the linear dimensions of head and pinnae. During development, the increasing interaural distance associated with a growing head and pinnae result in cues that change substantially and continuously until maturation is complete. These experiments examined whether the intersubject differences in the directional transfer functions (DTFs), the directional components of the head-related transfer functions, of cats during development could be reduced by scaling one set of DTFs on a log-frequency axis. Measurements were made of both the physical dimensions of the head and pinnae and the DTFs of cats aged 4 weeks to adulthood. For each pair of subjects, an optimal frequency scaling factor was computed that minimized the differences in the DTFs. In many cases, the optimal factors for pairs of subjects were proportional to the ratios of the measured physical dimensions of the head and pinnae in those subjects. Although there were still considerable differences in the DTFs even

after scaling, these data support frequency scaling as a potential method of modeling the development of the DTFs in cats. [Work supported by NIH Grant DC006865.]

3pSP11. A method to make a moving virtual sound source.
Ji-Ho Chang and Yang-Hann Kim (KAIST (Korea Adv. Inst. of Sci. and Technol.), 373-1 Kusong-dong, Yusong-gu, Daejeon 305-701, Korea)

This study deals with a method that manipulates the sound field, which is excited by a moving source. That is, the moving source is virtually constructed by array speakers. It is realized by controlling pressure values on boundary of the field. This is, in fact, an attempt to implement the Kirchhoff-Helmholtz integral equation in practice. It is well known that we can generate a certain sound field in a space by controlling pressure values on boundary [S. Ise, Acustica, **85**, 78–87 (1999)]. However, time varying sound field control has not been well addressed. Our aim is to manipulate sound field that is generated by a virtual source moving straightly outside the sound field or traveling around it. To express the method, the simple pure-tone case is considered and then expanded to more general cases. It also handled the problem related to eigenfrequencies at which boundary pressure control is not applicable. Computational and experimental results are introduced as well. [Work supported by BK21 & NRL.]

3pSP12. Partial-differential-equation-based sound source localization: Finite Fourier integral approach and its application to multiple source localization.
Yuya Fujita, Nobutaka Ono, and Shigeru Ando (Dept. of Information Phys. and Computing, Grad. School of Information Sci. and Technol., The Univ. of Tokyo, Tokyo, Japan)

In this paper, a method based on the sound source constraint PDE (partial differential equation) is developed for real complex environments with multiple sources. The sound source constraint PDE is integrated within a finite temporal interval with orthogonal Fourier bases to obtain simultaneous equations including source directions and distances as unknown variables. In them, the fast Fourier transform can be used rigorously and efficiently with least computation time. We show the equations enable us to obtain 3-D locations of up to four sources simultaneously from one spatial gradient observation of sound field. Numerical simulation comparison with the spatio-temporal-gradient-based method was carried out and showed the significant improvement of accuracy. Experimental results using a six proximity microphone array [N. Ono *et al.*, in Proc. ASJ Autumn meetingCvol.IC, pp. 607–608, September, 2004 (in Japanese)] in the real environment will be reported.

3pSP13. Holographic diagnosis: Basic principles and applications.
Jong-Hoon Jeon and Yang-Hann Kim (Ctr. for NOVIC, Dept. of Mech. Eng., KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea)

Acoustic signals achieved from a machine often have information on condition of the machine. This means that the location of faults can be found by using acoustic signal. Location of sound source can be found by using techniques such as MFAH (moving frame acoustic holography), which is applicable for pure tone noise and good signal-to-noise ratio, and beamforming, which basically assumes no reflected wave. Faults due to breakdown of rotating parts in a moving vehicle, however, make the above method hard to be directly applicable. The difficulty stems from three characteristics in the fault detection of the moving vehicle. First, the sound source radiates periodic impulsive noise. Second, there exist reflected waves from the ground since the array microphones should be settled between the vehicle and the ground to monitor the bottom side of the vehicle. Last, for the measurement problem, there exists noise due to other parts and wind for the moving vehicle, which degrades coherence among measured signals. In this research, holographic diagnosis, a machine diag-

nosis technique based on holographic approach, is introduced to find the position of faults effectively. The modified version of beamforming that models the moving periodic impulsive noise as a scan function is introduced.

3pSP14. Software-centered implementation of 128-channel loudspeaker array with stock PC and its application. Hiroshi Mizoguchi, Takafumi Nakamura, Hiroshi Takemura (Tokyo Univ. of Sci., 2641 Noda, Chiba 278-8510, Japan), and Satoshi Kagami (DHRC, AIST, Japan)

A huge speaker array system of 128 loudspeakers was constructed and experimented. It was implemented as "software-centered" style, utilizing stock loudspeakers and a PC. No dedicated hardware nor DSP was utilized. Spot forming, instead of beam forming, could be realized by 32 by 4 square layout of the array. Spot means small area of higher sound pressure level. Number of the spot was not limited to one. In the experiment, within 3 m by 3 m area, four spots of different sounds could be simultaneously formed. This spot forming was confirmed by actually measuring spatial distribution of sound pressure level. The effect of the spot was also confirmed auditorily. Since the system was software centered, it was dynamic. By simply changing software parameters, the location of the spot can be easily moved even while the system is running. This movability of the spot was intended to be the basis for visual steering. To realize the system, a simultaneous 128-channel 14-bit DA converter PCI board was developed. The 44.1-KHz sampling rate was achieved by a 2.4-GHz Intel Xeon based PC utilizing the DA board and a real-time OS, named ART-Linux. Approximately 23-μs loop could be realized by the software. It was the world's fastest software loop.

3pSP15. Complex envelope implementation for fast and efficient acoustic holography. Choon-Su Park and Yang-Hann Kim (Ctr. for Noise and Vib. Control, Dept. of Mech. Eng., KAIST, Daejeon 350-701, Republic of Korea)

Acoustic holography allows us to predict the spatial pressure distribution on any surface of interest. To implement this method, we normally use Fourier transform in time and space domain. It is noteworthy that the data size is so huge that it takes a long time to calculate pressure field. Moreover, the reconstructed pressure field is frequently too complex to observe the characteristics of pressure field. One possible candidate is the complex envelope. The complex envelope in time domain is well known and widely used in various fields. We have attempted to extend this method in the space domain, so that we can have a rather compact sound pressure picture that provides the information we need, for example where sound sources are. This belongs to what we call the analysis problem of acoustic holography. We might want to draw some parts in detail but other parts in a rather coarse way. The complex envelope in space certainly meets this objective. First we start with the simplest case. We devise a monopole complex envelope. Then we extend it to finite size source case. Various holography examples are reprocessed according to what we propose and demonstrate how this method is practically fast and how it provides a better picture for analyzing the sound field.

3pSP16. The acoustic information extraction of the indoor impulse response by multiresolution analysis. Wakako Tanaka, Yong Tang, Toru Itakura, and Hideo Shibayama (3-7-5 Toyosu Koutou-ku, Tokyo, 135-8548, Japan, m105072@sic.shibaura-it.ac.jp)

The impulse response is divided into approximation and detail components by the multiresolution analysis. The correlations between the early sound and reflected sounds are estimated respectively for the impulse response and the approximate signal. The approximate signal was calculated as a function of level of the multiresolution analysis. The correlation coefficient of these correlations is over 0.7 under level three. We evaluate the differences of these correlations from the listening test.

3pSP17. Detection for the early reflection in the impulse response by the correlation function. Toru Itakura, Yong Tang, Wakako Tanaka, and Hideo Shibayama (Shibaura Inst. of Technol., Toyosu 3-7-5, Koutoku-ku, Tokyo 135-0061, Japan, m106009@sic.shibaura-it.ac.jp)

Detecting an early reflection sound contained in the impulse response measured in a room, we calculated the correlation factor between the early reflection and the impulse response. Calculating the correlation factor every sampling period, we make the correlation function. For the function, we set a threshold level. We detected the absolute correlation value larger than the threshold level. And, we set the zero value for subthreshold level. Following this algorithm, we generate a new signal with the high correlation factor for the early reflection. We compared the signal with the impulse response by using listening test.

3pSP18. A design of virtual sound field simulator based on acoustic ray tracing and wall impedance. Kiichiro Yamada, Yuki Denda, Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, rs047023@se.ritsumei.ac.jp), and Takanobu Nishiura (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

The sound propagation analysis using an architectural scale model is an effective approach for architectural acoustic design of a concert hall, and so on. However, the scale model must be reconstructed together, changing the basic plan of the architectural design. To cope with this problem, an architectural sound field has been virtually designed on the computer by the sound field simulator in recent years. The acoustic ray tracing is one of the simulators based on geometric acoustics. It is an efficient solution from the viewpoints of simulation cost and time. However, it cannot simulate the wall materials, because it employs the delta function as the wall impedance. To overcome this problem, in this paper, we try to design the virtual sound field simulator based on acoustic ray tracing and wall impedance acquired by real measurements. We first measured the reflection characteristics of the wall with impulse response to acquire the wall impedance. Then, we designed the virtual sound field simulator by convoluting the acoustic rays with measured wall impedances. As a result of subjective evaluation experiments, the proposed simulator could realize a more realistic sound field than the conventional acoustic ray tracing. [Work supported by MEXT of Japan.]

3pSP19. A fundamental investigation of the small microphone array system and its applications. Akira Nawa, Yoku Watanabe, and Hareo Hamada (Tokyo Denki Univ., 2-1200 Muzai-gakuendai, Inzai, Choba prefecture 270-1382, Japan)

A purpose of this research is to construct a 3-D sound recording and reproduction system that has a high presence by performing a spatially controlled sound field synthesis technique. For this purpose, it will be required to examine and choose the most suitable sound recording system, since a conventional sound recording system samples only original signal itself, and does not include enough spatial information. In other words, it will be required to examine the system to record the information of the three-dimensional spaces, such as the direction and the distance, and to transform the configuration to become effective to the reproduction system. This paper will deal with a basic examination to construct the microphone array placed in proximity and the recording system, which is able to control the directivities in the three-dimensional spaces, using algorithm based on LMS by FIR filter networks from signals to each microphone.

3p THU. PM

3pSP20. Blind source separation of musical signals applied to selectable-listening-point audio reconstruction. Kenta Niwa (Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya-city, Aichi, Japan, niwa@sp.m.is.nagoya-u.ac.jp), Takanori Nishino, Chiyomi Miyajima, and Kazuya Takeda (Nagoya Univ., Japan)

An audio reproduction system is being developed to generate binaural signals at user-selected listening positions from audio signals of multiple sound sources recorded with multiple microphones. The system consists of a sound decomposer based on blind source separation (BSS) and a sound recomposer using head-related transfer functions (HRTFs). In this work, musical signals of multiple instruments are separated into the original independent signals of the respective instruments using BSS based on frequency-domain independent component analysis (FD-ICA). HRTFs are then convoluted to the separated sound sources to reconstruct the music sounds of the selected listening point. Preliminary experiments are conducted using two or three sound sources of instruments: flute, violin, and piano generated by a synthesizer. The synthesizer sounds are played through small directional loudspeakers or spherical omnidirectional loudspeakers in a low-reverberant sound-proof room or moderately reverberant lecture room and recorded with multiple microphones. Live music played on real instruments including flute, violin, and trumpet are also recorded in the same rooms. Audio signals of respective instruments obtained by BSS are available on our web site.

3pSP21. Acoustic sound field simulator based on head-related transfer function and three-dimensional mirror-image. Keitaro Kuroda, Yuki Denda, Masato Nakayama (Grad. School of Sci. and Eng., Ritsumeikan Univ., 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan, rs016029@se.ritsumei.ac.jp), and Takanobu Nishiura (Ritsumeikan Univ., Kusatsu, Shiga 525-8577, Japan)

The head-related transfer function (HRTF) is the most important factor to simulate an acoustic sound field for the virtual reality technology or the architecture field. However, the HRTF has individual differences on every listener, because it represents the real transfer function between the sound source point in the acoustic field and the listener's ears. In addition, it is impossible to measure the HRTFs on every listener. To solve this problem, we focus on three-dimensional mirror-image. It can represent the virtual transfer function (especially, room reverberation) in an acoustic field by replacing all surface reflections with mirror-image sources. Therefore, we try to design the new acoustic sound field simulator by convoluting the real transfer function of the HRTF and the virtual transfer functions of the three-dimensional mirror-image. By this approach, the proposed simulator may realize the more accurate sound-image localization with room reverberation by three-dimensional mirror-image even if the HRTF of an other listener is employed. As a result of evaluation experiments in some conditions, the proposed simulator could improve the sound-image localization performance especially in an anteroposterior direction more than the conventional simulator based on HRTF. [Work supported by MEXT of Japan.]

3pSP22. Binaural transformation coding with simulated head tracking. Seiichiro Shoji and Anthony I. Tew (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD UK)

The binaural transformation codec synthesizes generic binaural audio signals and generates accompanying side information. In the decoder the side information is used to personalize the generic signal. The performance of the basic codec is described in another paper; this paper describes the estimated effect on quality of incorporating limited head tracking. Generic HRTFs were used to spatialize two concurrent sound sources spatialized at ±30 or ±80 deg. The generic binaural signal was personalized in the BX decoder and at the same time the individual sound sources were rotated. These induced rotations are equivalent to compensating for head yaw rotations of −80, −40, −20, −10, and −5 deg, and for head pitch rotations of 22.5, 45, and 90 deg. Listening tests based on Recommendation ITU-R BS.1116-1 were used to evaluate the processed

sounds. The tests were conducted using speech, vocals, guitars, and percussion source materials. It was found that the quality of the processed sound tended to degrade as the head yaw rotation increased. However, except for the head yaw condition −80 deg, the quality of the processed sound remained greater than 4.0 on the mean opinion scale. No deterioration of the processed sound was observed for the applied pitch rotations.

3pSP23. Novel method for the efficient individualization of binaural audio signals. Seiichiro Shoji and Anthony I. Tew (Electron. Dept., Univ. of York, Heslington, York, YO10 5DD UK)

A novel coding technique suitable for the transmission of binaural audio signals with concurrent multiple sound sources is proposed in this paper. Existing algorithms typically suffer from audible artefacts, poor localization, or excessive bandwidth. Especially with conventional intra-source coding methods, bandwidth is dependent on the number of sources and becomes greater as the number of sources increases. The proposed binaural transformation (BX) encoder synthesizes binaural signals using generic HRTFs and transmits them to the decoder along with side information. The bandwidth of the side information is independent of the number of sources. Using locally stored HRTFs matching those of the listener, the decoder turns the generic binaural signals into personalized ones. We present two types of BX codec: type A and type B. Psychoacoustic experiments based on Recommendation ITU-R BS.1116-1 were used to compare signals processed by these codecs with the corresponding unprocessed signals. For a variety of source materials (speech, vocals, guitars, and percussion) we show that high mean opinion scores are achievable using the type A codec when its parameters are set appropriately.

3pSP24. Individuality of elevation-angle-sensitive notches in head-related transfer functions. Saho Sekimoto, Yukio Iwaya, and Yôiti Suzuki (Res. Inst. of Elec. Commun./Grad.. School of Information Sci., Tohoku Univ. Katahira 2-1-1 Aoba-ku, Sendai, 980-8577, Japan, saho@ais.riec.tohoku.ac.jp)

Perceived localization can be controlled with high accuracy in the rendered auditory space when listeners' own head-related transfer functions (HRTFs) are used for rendering 3-D spatial audio. However, measurement of a listener's own HRTFs in all directions requires a special measuring apparatus and a long measurement time, which entail a considerable physical load on subjects. Therefore, effective individualization methods of HRTFs must be very useful. In previous studies, a few notches around 6–10 kHz, whose frequencies change depending on the elevation, are inferred as important to determine the perceived elevation [e.g., E. A. G. Shaw and R. Teranishi, J. Acoust. Soc. Am. **44**, 240–249 (1968)]. Therefore, this study examines the individuality of such notches, of which the frequency is elevation-angle sensitive. First, individual HRTFs from all directions are measured for 140 listeners. They are then analyzed using an autoregressive moving average (ARMA) model. Second, the frequencies of these notches are expressed smoothly as a function of the elevation angle using dynamic programming (DP). Results qualitatively clarified the dependency of each frequency on elevation. Moreover, distributions of the notch frequencies show systematic variation as the elevation angle changes.

3pSP25. Objective and subjective evaluation of numerically estimated head-related transfer functions. Shouichi Takane and Koji Abe (Dept. of Electron. and Information Systems, Akita PRef. Univ., 84-4 Ebinokuchi, Tsuchiya, Yurihonjo, Akita, 015-0055 Japan, takane@akita-pu.ac.jp)

In this study, head-related transfer functions (HRTFs) of a dummy head and human subjects are numerically estimated by using the boundary element method (BEM), and their availability for auralization is discussed from the subjective viewpoints. Geometries of their heads are obtained by a 3-D laser scanner. The effect of the torso for the estimation of the HRTFs is also examined and discussed via the 3-D measurement. The estimated

HRTFs are compared for each subject with the measured HRTFs in various objective criteria, such as spectral distortion (SD), signal-to-distortion ratio (SDR), and interaural time/level differences (ITD/ILD). Availability of the estimated HRTFs for the synthesis of the 3-D sound image are then evaluated via hearing experiments. In our previous study [Proceedings of WESPAC IX (2006)], the results of the experiment is insufficient since the low-frequency range of the estimated HRTFs caused the ambiguity in sound localization. In this study, the frequency range of the estimation is extended to as high frequency as possible, and the availability of the estimated HRTFs for auralization is discussed from two viewpoints: the subjective difference between the estimated and measured HRTFs, and some subjective attributes obtained from the hearing of the stimuli convoluted with the estimated HRTFs.

3pSP26. Analysis and modeling of interaction of auditory-visual information. Takahumi Asada, Yuko Watanabe, and Tatsuya Shibata (Grad. School of Information Environment, Tokyo Denki Univ., 2-1200 Muzai Gakuendai Inzai, Chiba, 270-1382 Japan, den03117@nifty.com)

We analyze and model the interaction of movement of the synchronized auditory-visual information. We construct the relationship of movement between virtually sound source and object. We suppose that cubic local contrast model (CLCM) we propose and the optical flow method are useful for analyzing moving image. CLCM is a method that describes the movement of objects from temporal and spatial difference of local contrast between pixels in moving image, and we found CLCM can measure the orientation of camera motion and the moving direction of an object in a moving image. We use the feature of head-related transfer function (HRTF) as the feature of sound movement. We have done two experiments: experiment 1 is that we analyze the data of the interaural amplitude difference (IAD) and interaural phase difference (IPD) between right and left HRTFs of a sphere by local contrast model (LCM); experiment 2 is that we analyze the relationship between the movement of an object in a moving image and a virtual sound source. We discuss the interactive features between auditory and visual information.

3pSP27. The effect of head movement on sound localization using stereo microphones: Comparison with dummy heads. Wataru Endo (Tokyo Inst. Tech., Nagatsuda, Midori-ku, Yokohama 226-8502, Japan, endo@u.ip.titech.ac.jp), Makio Kashino (Tokyo Inst. Tech., Atsugi, Kanagawa 243-0198, Japan), and Tatsuya Hirahara (Toyama Prefectural Univ., Imizu, Toyama 939-0398, Japan)

This study examined the effect of head movement on sound localization with a pair of microphones providing no head-related transfer function (HRTF) information, and with individualized, nonindividualized, and downsized dummy heads providing HRTF information with different degrees of distortion. In an anechoic room, white noise was presented for 5 s from one of 12 loudspeakers. A dummy head or a microphone pair was placed at the observation point. Listeners heard the sound via headphones outside the anechoic room and judged the direction of the source and whether the sound image was extracranial. Without head movement, localization accuracy with the microphone pair was significantly worse than that with dummy heads. When voluntary head movement was allowed and the dummy head was moved in accord with the listener's movement, accuracy with nonindividualized and downsized dummy heads improved nearly to the individualized level. Accuracy with the microphone pair also improved with head movement, but did not reach the level of dummy heads. An extracranial image was perceived with dummy heads in most cases, but not often with the microphone pair even with head movement. Head movement cannot completely compensate for the lack of HRTFs.

3pSP28. A biologically inspired pitch determination algorithm. Arturo Camacho and John G. Harris (Computational Neuro-Eng. Lab, Elec. and Comput. Eng., Univ. of Florida, Gainesville, FL 32611, arturo@cnel.ufl.edu)

A biologically inspired pitch determination algorithm is presented. This algorithm combines existing models of the cochlea and inner-hair-cell based spike generation [Lopez *et al.*, J. Acoust. Soc. Am. **110**, 3107 (2001); Sumner *et al.*, *ibid.* **113**, 893–901 (2003)] to model spike trains in the auditory nerve. The pitch and its salience are then estimated using a method proposed by Cariani [Cariani *et al.*, J. Neurophysiol. **76**, 1698–1716], which computes a summary autocorrelation function over the spike trains. Unlike Cariani's work, where spike trains are obtained experimentally, we simulate the spike trains from biologically inspired models. The main contribution of our approach is to combine models of the auditory system and a pitch estimation method based on neural spike trains. The proposed algorithm was tested using standard synthesized sounds, speech, and singing voices. Results show that this algorithm better matches human performance as compared to traditional pitch detection algorithms used in automatic speech processing.

3pSP29. End-point detection of speech using spectral transition for captioning system. Ayako Koga, Yuki Fujikashi, Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554), Noboru Kanedera (Ishikawa Natl. College of Technol.), and Junko Yoshii (Fujiyama, Inc.)

In recent years, captioning video contents with text translations is increasingly necessary because of the burgeoning use of media internationally, resulting from the rapid development of communication technology. In addition, within one language, video captioning is very important for hearing-impaired people. However, the process of captioning videos, including speech and nonspeech decisions, is often done manually by translators at present. Therefore, an efficient automatic end-point detection of speech for captioning video contents has been proposed. We attempted to detect speech end-points based on acoustic landmarks that identify times when acoustic changes are prominent in the speech signals [K. N. Stevens, Acoust. Phonetics (1998)]. In this study, landmarks were obtained by combining the mean square for the regression coefficients of logarithmic envelopes of $\frac{1}{3}$-oct bands in time, which resembles the parameter proposed by Furui to measure spectral transition [S. Furui, J. Acoust. Soc. Am. **80**(4), 1016–1025 (1986)], with other ones such as the logarithmic power of speech signals. An experiment was carried out using the proposed technique for speech detection. Results showed a high correct rate and introduced the possibility of its application to an efficient video captioning system. [Work partially supported by Open Research Center Project from MEXT.]

3pSP30. Inverse filter analysis of common harmonic structure on Specmurt using Riemann's ζ function. Nobutaka Ono, Shoichiro Saito, Hirokazu Kameoka, and Shigeki Sagayama (Dept. Information Phys. and Computing, Grad. School of Information Sci. and Technol., Univ. Tokyo, 7-3-1 Hongo Bunkyo-ku Tokyo, 113-8656 Japan)

In this paper, based on the interesting relationship between the log-spaced δ function sequence and Riemann's ζ function, the analytical properties of the inverse filter of the common harmonic structure on *specmurt analysis* are discussed. Specmurt [S. Sagayama *et. al.*, in Proc. SAPA (2004)] is a simple and efficient technique for the multi-pitch analysis of polyphonic music signals. If all tones have the same harmonic pattern, the power spectrum on the log-scaled frequency can be regarded as the convolution of the common harmonic structure and the distribution of fundamental frequencies. Based on the model, overtones are effectively suppressed by the inverse filtering of the common harmonic structure in specmurt. Thus, for the stable processing, analytic properties of the inverse filter are significant. Our new finding is that when the common harmonic structure is expressed as a log-spaced δ function sequence with a particular kind of decay, the Fourier transform is exactly equal to Riemann's ζ

function. Interpreting several properties of Riemann's ζ function in a signal processing context gives us the new perspectives like an explicit representation or cascaded decomposition of the inverse filter.

3pSP31. The calculation of sleep apnea syndorome's condition index using the snore noise. Takahiko Ono (Tokyo Univ. of Agriculture and Technol., 3-8-1, Harumicho, Fuchu, Japan), HIroo Yano, Tatsuya Oota (Chiba Inst. of Technol., Chiba, Japan), Yuuji Takasaki, and Sachiko Kurosaki (Oota Kinen Hospital, Kooriyama, Japan)

The polysomnograph analysis is used to measure an apnea hypopnea index (AHI) of sleep apnea syndorome. However, this method needs high grade devices and technique. To compensate the conventional method, the calculation of AHI using only the snore noise of the patient is invented.

3pSP32. Voice activity detection with voice/nonvoice classification using reliable fundamental frequency. Soo-young Suk, Hiroaki Kojima (Information Technol. Res. Inst., AIST, Central 2, Umezono, Tsukuba, Japan, sy.suk@aist.go.jp), and Hyun-Yeol Chung (Yeungnam Univ.)

In this paper, we introduce a voice activity detection (VAD) algorithm to perform voice/nonvoice (V/NV) classification using a fundamental frequency ($F0$) estimator called YIN. Although current speech recognition technology has achieved high performance, it is insufficient for some applications where high reliability is required, such as voice control of powered-wheelchairs for handicapped persons. The proposed VAD, which rejects nonvoice input in preprocessing, is helpful for realizing a highly reliable system. Previous V/NV classification algorithms have generally adopted statistical analyses of $F0$, the zero-crossing rate, and the energy of short-time segments. A combination of these methods, a cepstrum-based $F0$ extractor, has been proposed [S. Ahmadi and S. S. Andreas, IEEE Trans. SAP. **7**, 333–339 (1999)]. The proposed V/NV classification adopts the ratio of a reliable fundamental frequency contour to the whole input interval. To evaluate the performance of our proposed method, we used 1360 voice commands and 736 noises in powered-wheelchair control in a real environment. These results indicate that the recall rate is 97.4% when the lowest threshold is selected for noise classification with precision 97.3% in VAD. The proposed VAD, which rejects nonvoice input in preprocessing, can be helpful to realize a highly reliable system.

3pSP33. Estimation of a talker and listener's positions in a car using binaural signals. Madoka Takimoto (Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8603 Japan, takimoto@sp.m.is.nagoya-u.ac.jp), Takanori Nishino, and Kazuya Takeda (Nagoya Univ., Japan)

The problem of estimating the location of a sound source is described which is based on signals observed at the entrances of the two ears. The purpose is to specify the talker's and listener's positions within a car using the binaural signal. The talker and the listener sit in two of the four car seats. In this experiment, two kinds of head and torso simulators are used as a talker and a listener. Given information includes the acoustic transfer functions for all positional patterns. Eight patterns of acoustic transfer functions are measured, involving those that have the same positional pattern, but in which the talker faces a different direction. A Gaussian mixture model for each positional pattern is generated. The parameters we used are interaural information such as the envelope of an interaural level difference. The models are evaluated by specifying the positional pattern. Results show that we can specify positions with up to 97% (35/36) accuracy using the binaural signals of two men. Then the input signal was expanded to one with background noise that resembles a real situation, and a model that involves motion of the talker's head was also considered.

3pSP34. An implementation of realtime speech speed control system using hierarchical cascaded integrator and comb filters. Hironobu Takei, Shogo Nakamura, Hiroto Saitoh, and I. Nyoman Sukadana (Tokyo Denki Univ., 2-1200 Muzai Gakuendai, Inzai Chiba, Japan, takei@sp.sie.dendai.ac.jp)

In this paper, our algorithm for speech speed control system and its DSP implementation are described. The system slows or speeds up speech without changing its pitch, which aids learning a foreign language, listening for older people, and so on. The proposed algorithm is based on the phase vocoder, which doesn't need extraction of voiced fragments from a speech signal. And the reconstruction of time-scaled signal involves no smoothing operation between voiced and unvoiced sections. The speech signal is decomposed to bandpassed frequency channels so that the envelope of each of the channels is stretched. Then the stretched envelope is multiplied by the corresponding sinusoidal, and by summing up the channel components the speed controlled signal is reconstructed. Each band has very narrow bandwidth, which requires a large amount of computation with a FIR-based design. To achieve realtime operation of our system, a hierarchical structured cascaded integrator and comb (HSCIC) filter is also introduced, which can have highly sharpened characteristics with only a few multipliers.

3pSP35. On the accuracy of speech source direction identification on the horizontal plane using 3-D audio displays. Hirotoshi Terada, Kazuhiro Kondo, and Kiyoshi Nakagawa (Yamagata Univ., Jounan 4-3-16, Yonazawa, 992-8510, Yamagata, Japan)

This paper describes initial results of speech source direction identification experiments on the horizontal plane. This is an initial study for the sound localization required in order to separate the individual speech source location to increase the intelligibility in remote conference systems. Our focus here was to determine the effect of sound source characteristics on the source direction identification accuracy. Sound sources were located on three target locations (forward, right, and left) and five (45 deg in both right and left added) on the horizontal plane using a real-time virtual acoustic environment rendering system (NASA SLAB). SLAB integrates HRTFs to render the source localization. The sound sources used were conversation speech, white noise, and conversation speech mixed with noise at an SNR of 0 dB. The observers were asked to identify the direction of the rendered sound presented over a headphone. The results showed that the direction identification accuracy of conversational speech is approximately 20% higher than that of white noise in all directions tested. However, the accuracy of mixed conversation and white noise was about the same as the accuracy with white noise only, suggesting that the high-frequency components may be degrading the perceived accuracy.

3pSP36. Multipitch estimation based on correlation between a harmonic structure model and an observed spectrum for mixed audio signals. Yoichi Uchida and Naotoshi Osaka (Tokyo Denki Univ., School of Eng., 2-2 Kanda Nishiki-cho, Chiyoda-ku, Tokyo, 101-8457, Japan, uchida@srl.im.dendai.ac.jp)

In the research of sound source separation and resynthesis, multipitch estimation is an important topic. We propose a multipitch estimation technique for monaural audio signals. It introduces a method of calculating multiple pitches within a frame for a pitched sound signal. First, a harmonic structure is modeled for a pitch M (Hz), using the cases of odd integer multiples of M, even integer multiples of M, and odd integer multiples of $M/2$. Then correlations are derived between each of the harmonic structure models and an observed spectrum. Harmonic structure models are expressed as the summation of Gaussian functions in the frequency domain, where each harmonic has a peak at its center frequency. An evaluation function is dynamically defined to test how well arbitrary values of M fit the observed spectrum. The evaluation function is created by searching for weighted coefficients for the correlation values. Next, a number of maximal peaks equal to the number of sound sources are se-

lected from the spectrum and defined as the estimation pitches. This method was applied to a mixed piano signal, and a mixed audio signal that includes flute, clarinet, and cello. The results are good except at octave unisons.

3pSP37. Transient detection algorithms for speech coding applications. Grzegorz Szwoch, Maciej Kulesza, and Andrzej Czyzewski (Gdansk Univ. of Technol., Narutowicza 11/12, 80-952 Gdansk, Poland)

One of the methods for improving transmitted speech quality is efficient encoding of transients. The majority of currently used speech codecs employ the parametric approach, thus they do not differentiate between transients and other signal components. A sine+noise+transients signal model, proposed earlier as a tool for wideband audio signal encoding, seems to be a promising solution for enhancing signal quality in speech coding. Methods of transient identifying for an efficient transform-domain encoding of transients has to be chosen carefully. Different transient detection algorithms are described in literature, most of them optimized for musical signals. The methods do not ensure efficient transient detection in speech signal, due to the different character of musical signals and speech. This paper provides a comparison of several transient detection methods, mostly based on evaluation of signal energy, that are now primarily used for musical signals. Their efficiency and application to transient detection in speech are evaluated and discussed. Differences between transient detection in wideband musical signals and speech are pointed out. Modifications to existing algorithms are proposed, allowing efficient transient detection and encoding in the proposed, wideband speech codec. [Research funded by the Polish Ministry of Education and Science within Grant No. 3 T11D 004 28.]

3pSP38. Sound source classification using temporal pattern for mixed sounds. Toru Taniguchi, Katsuhiko Shirai (Dept. of Comput. Sci., Waseda Univ., 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555 Japan, tani@shirai.cs.waseda.ac.jp), and Mikio Tohyama (Waseda Univ., Honjo-shi, Saitama 367-0035, Japan)

In speech information processing of real-world environments, how to deal with nonstationary acoustic noise such as background music or speech is a remaining problem while much progress has been made in the speech processing in stationary noise. Music or speech is a sound that has a complex spectral and temporal structure composed of many musical notes or phonemes, and this fact makes it difficult to estimate signals of music or speech separated from a foreground speech. For noise estimation of one-channel sound, only the prior knowledge of temporal or spectral structures of the noises is available besides the signal itself. Therefore, dealing with nonstationary noise of music or speech essentially requires the models representing the temporal and spectral structure of them. In our study, a classification method of nonstationary sounds using sinusoidal decomposition is proposed. In the method a classification of mixed sounds into sound categories, like speech, singing voices, or instruments, is performed based on temporal characteristics of spectral sequences analyzed by a spectral peak tracking algorithm. Temporal features are inspected by statistical analyses and discrimination experiments of sinusoidal sequences. The classification algorithm is evaluated by experiments of classification of mixed sound.

3pSP39. Multiple fundamental frequency estimation assisted by audio watermarking. Yi-Wen Liu (Boys Town Natl. Res. Hospital, 555 North 30th St., Omaha, NE 68131)

The success of computer segregation of sound sources from a single-channel mixture often relies on the estimation of multiple fundamental frequencies. Instead of solving the problem directly, this presented work utilizes audio watermarks to assist sound source segregation. Here, individual sources are assumed to be available for watermark embedding be-

fore mixing. Each source's short-time spectral peaks are aligned to frequency gridpoints labeled with a digital ID [Liu and Smith, "Audio Watermarking Based on Sinusoidal Analysis and Synthesis," International Symposium on Musical Acoustics, Nara, Japan (2004)]. A modified sound source is sinusoidally synthesized afterwards. In this sense, each source is embedded with a watermark. In an experiment, two sound sources are watermarked with two sets of gridpoints P and Q, respectively. Points in P and Q interleave and are spaced 10 cents of a semitone apart. The watermarked sources are then linearly added. To demix them, sinusoidal trajectories found in the mixture are segregated based on whether its frequency more often aligns to P or Q. Then, segregated signals are reconstructed by sinusoidal synthesis. Although the reconstructed signals do not sound like the original sources, the melodies of the original are successfully extracted from the reconstructed in monophonic pitch estimation.

3pSP40. Directions of arrival (DOA) estimation by particle filtering with dominant frequency components in noisy environments. Mitsunori Mizumachi, Norikazu Ikoma, and Katsuyuki Niyada (Kyushu Inst. of Technol., 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan)

In noisy environments, DOA estimation is useful for achieving attractive human-machine interfaces in the real world. This article proposes a robust DOA finder based on cross correlations between observed signals by particle filtering with frequency selectivity. The particle filter is flexible for modeling the system because it requires no linearity or Gaussianity on the model. The temporal trajectory of DOAs is modeled using a Markovian state space model. It is estimated through the state estimation procedure using particle filters. Markov modeling is expected to achieve stable DOA estimation in time. It is also a key for the proposed method to selectively use the dominant frequency components, such as formant frequencies of speech, to improve noise robustness. Assuming the quasi-stationarity of noise signals in the presence of a time-varying target speech signal, the dominant frequencies are extracted using a deterministic approach similarly to spectral subtraction. A set of the subband cross-correlation functions at the dominant frequencies is regarded as likelihood and is used for updating particle weights to estimate posterior distribution. The feasibility of introducing both particle filtering and frequency selectivity into DOA estimation was confirmed in several noise conditions.

3pSP41. Azimuth and elevation estimation for multiple sound sources using two-dimensional frequency domain binaural model. Yoshifumi Chisaki, Sho Kawano, and Tsuyoshi Usagawa (Grad. School of Sci. and Technol., Kumamoto Univ., 2-39-1, Kurokami, Kumamoto, 860-8555 Japan, chisaki@cs.kumamoto-u.ac.jp)

Direction of arrival (DOA) estimation for multiple sound sources with only two sensors as the human being does is proposed. The frequency domain binaural model (FDBM) [Nakashima et al., Acoust. Sci. Technol. **24**, 172–178 (2003)] was proposed as one binaural model. The method uses sound diffraction to estimate DOAs and segregate sound sources. The binaural model can estimate one-dimensional DOA, namely, azimuth. However, human beings can estimate not only azimuth of DOA but also elevation of that. In order to achieve an estimate of both azimuth and elevation simultaneously, a new DOA estimation algorithm based on interaural phase and level differences is proposed as a part of frequency domain binaural model. Results of DOA characteristics of the proposed binaural model estimation on multiple sound sources' condition is compared with that of human auditory characteristics. As a result, it was confirmed that performance on accuracy of DOA estimation depends on power level ratio among sound sources. [Research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C), 2006, 18500135 and carried out by the Cooperative Research Project Program of the RIEC, Tohoku University (Grant No. H16.A02).]

3pSP42. Sound source localization using a distributed microphone system in real environments. Kook Cho, Hajime Okumura, Takanobu Nishiura, and Yoichi Yamashita (College of Information Sci. and Eng., Univ. of Ritsumeikan, 1-1-1, Noji-Higashi, Kusatsu-shi, Shiga 525-8577 Japan)

Accurate localization of sound sources is indispensable for high-quality sound capture. The difference of distance from a sound source to two microphones causes a time delay of arrival (TDOA). The cross-power spectrum phase analysis (CSP) method has been proposed to estimate the TDOA based on correlation in signals. The sound source can be localized theoretically as a crossing point of sound directions estimated using two pairs of microphones. However, in a real environment, microphones receive sounds that have originated through reverberation and reflection. This situation makes TDOA estimation difficult, and accurate sound localization is difficult when using only two pairs of microphones. The authors propose a new method of sound localization based on multiple microphones that are widely distributed and placed under the ceiling of a room. A number of pairs of microphones give a series of CSP correlation coefficients as a function of the time delay. The CSP coefficients for hypothetical sound sources are accumulated over all the microphone pairs. The number of microphones is 16. The sound source is localized as a hypothesis that maximizes the accumulated CSP coefficient. Experiment results obtained in a real room show that the proposed method improves the localization accuracy.

3pSP43. Free listening-point synthesizing method in a large microphone array using acoustic transfer function. Mehrdad Panahpour Tehrani, Yasushi Hirano, Shoji Kajita, Kenji Mase (Information Technol. Ctr., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan), Takanori Nishino, Kazuya Takeda (Grad. School of Information Sci., Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan), and Toshiaki Fujii (Grad. School of Eng., Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan)

In this research a free listening-point synthesizing method is proposed by using the measured acoustic transfer function (ATF) in a large microphone array for future 3-D audiovisual systems. ATFs at the location of far-field microphones are estimated in frequency domain by using a recorded time-stretched pulse (TSP) signal broadcasted by speakers at the location of sources. Then an ATF pool is generated with a chosen density by synthesizing the estimated ATFs for each sound source. The ATF pool contains all the microphones' locations, plus synthesized ATFs with a chosen density from sound sources to any point on the arc which passes all of the microphones. The free listening-point is generated by selecting the ATF for each source from the ATF pool. Then the source microphone sounds are convolved with the chosen ATFs and added. The proposed method is evaluated by an array with 38 microphones in the far field on an arc with $r = 4$ m with 20 cm between each pair of microphones, and two microphones close to two sound sources with 1.5-m distance. Experimental results have been shown for the estimated ATFs and comparison of the synthesized arbitrary listening point with the actually recorded sound for a given point that shows satisfactory results.

3pSP44. On the numerical simulation of sound wave propagation using field programmable gate array device. Takao Tsuchiya (Dept. of Information Systems Design, Doshisha Univ., 1-3 Miyakodani, Tatara, Kyotanabe-shi, Kyoto 610-0321, Japan, ttsuchiy@mail.doshisha.ac.jp), Eiko Sugawara, and Yasusi Inoguchi (Japan Adv. Inst. of Sci. and Technol., 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan)

Numerical simulation of sound wave propagation using field programmable gate array (FPGA) device is proposed. The two-dimensional digital Huygensf model (DHM) or the transmission line matrix (TLM) model is configured in FPGA device as digital circuits. DHM is a numerical model in which the propagation and scattering of waves are simulated as the sequences of impulses scattering on the transmission line network as Huygens' principle states. The digital equivalent circuit can be easily devel-

oped based on DHM because the formulation of DHM is very simple compared with the finite difference time domain (FDTD) method. In this paper, number of logic cells of the FPGA device required for a DHM element is estimated. To configure DHM elements in the FPGA device as much as possible, the fixed-point operations are applied to the DHM calculations. Numerical precision of DHM using the fixed-point operations is then discussed. The data length required for acoustic field simulation in practical precision is estimated. The capability of the FPGA simulator for room acoustics is discussed.

3pSP45. Signal processing for small-size microphone array system— Application to teleconferencing system. Akira Gotoh, Yoshitaka Murayama, Jun Ota, Hareo Hamada (DiMAGIC, Sugishou Bldg 3F, 2-4-5 Kanda-Ogawamachi, Chiyodaku, Tokyo 101-0052, Japan), Yuko Watanabe, and Hareo Hamada (Tokyo Denki Univ., 2-1200 Muzaigakuendai, Inzai-shi, Chiba, Japan)

A small microphone array system, consisting of two or more closely arranged small microphones and a digital filter network, is newly proposed. A method based on sound field control using multiple-point least-squares filter design contributed to the configuration of a smaller size microphone array system. Several directivity patterns having unique focus directions can be configured by changing each set of optimum filter coefficients. In this paper, a new method of scanning speaker direction using the small microphone array is proposed and several performances used in teleconferencing systems are evaluated.

3pSP46. Sensor signal processing circuit for echolocation using a delta-sigma modulated single-bit digital signal. Shinnosuke Hirata, Minoru K. Kurosawa (Tokyo Inst. of Technol., G2-32, 4259 Nagatsuta, Midori-ku, Yokohama, Japan, shinnosuke.hirata@ip.titech.ac.jp), and Takashi Katagiri (Sutekina Inc. Komagane, Japan)

Ultrasonic sensors are suitable for perception of the environment of mobile autonomous robots. Most bats use echolocation that is a high-quality sensing system by their ultrasonic sensors [S. J. Gaioni, H. Riquimaroux, and N. Suga, "Biosonar behavior of mustached bats swung on a pendulum prior to cortical ablation," J. Neurophysiol. **64**(6), 1801–1817 (1990)]. The echolocation can estimate the distance to the object, its relative velocity, and its scale. But, technological application of echolocation necessitates a large computational requirement because delay of the echo signal from the generated signal is estimated from cross correlation between the generated signal and the echo signal. A new sensor signal processing for echolocation is proposed by using a delta-sigma modulated single-bit digital signal. A delta-sigma modulated single-bit digital signal is a single-bit high-speed data stream. [S.R. Norsworthy, R. Schreier, and G. C. Temes, *Delta-Sigma Data Converters* (IEEE Press, New York, 1997] Using delta-sigma modulated single-bit digital signal processing, multiplication and accumulation of cross-correlation can be changed into logical operations: EXOR, NOT, SUM. Simulations show the possibility of estimation of the distance to the object from correlation of the echo signal with the generated signal converted single-bit data stream, along with reduced computational and circuit-scale requirements.

3pSP47. Modeling of phased antenna array of about 500 microphones, detecting landing airplanes. Gaetano Caronna, Ivan Roselli, and Pierluigi Testa (Dipartimento di Fisica Tecnica, Universit La Sapienza, Via Eudossiana, 18, 00100 Roma, Italy, gaetano.caronna@uniroma1.it)

The University La Sapienza participated recently in the project SAFE-AIRPORT funded within the sixth framework program of the European Commission. This project aimed at developing a phased antenna array of about 500 microphones in order to acoustically detect and track flying airplanes. A preliminary field testing was performed at Rome-Fiumicino airport, however experimental data are not yet fully available. For this

reason, a computer-aided model of the antenna phased array was developed and beam-forming was simulated with real input signals (acoustic emissions from a real landing aircraft with superimposed disturbing, spatially uncorrelated, pink noise of realistic level). Detection range was measured by applying classic shift and sum beamforming. An algorithm was tested in order to increase maximum detection range by exploiting information from more than one temporal frequency, assuming that disturbing environmental noise is uncorrelated on different temporal frequencies. Specific results are reported.

3pSP48. Compressive coding for multichannel audio signals using independent component analysis. Tadashi Mihashi, Tomoya Takatani, Shigeki Miyabe, Yoshimitsu Mori, Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma-shi, Nara 630-0192, Japan)

This study addresses a novel compressive coding for multichannel audio signals with independent component analysis (ICA). In recent years, multichannel audio contents are widely spreading, and thus the transmission quantity is increasing. Consequently, audio-signal compression is a problem demanding prompt attention. Conventional compression methods utilizing sparseness assumption among audio-source signals have a problem that the method is likely to distort original sources owing to the lack of sparseness in the mixed signals consisting of two or more sound sources. To solve this problem, we propose an audio-signal compression strategy using ICA; ICA assumes that the source signals are mutually independent, and can estimate the source signals from the mixed signals without *a priori* information by applying an unmixing matrix. The decomposition of mixed signals leads to an enhancement of sparseness among signals. Furthermore, we propose a mechanism to detect time-frequency grids where sparseness is not satisfied, and for the detected grids we directly transmit multichannel signals to mitigate the distortion. The reconstruction process utilizes an inverse projection of the unmixing matrix to make correct lateralization of each sound source. Experimental result shows that the proposed method can compress mixed audio signals more efficiently and achieve a higher quality than the conventional method.

3pSP49. Novel interface using a microphone array for wheelchair control. Akira Sasou and Hiroaki Kojima (Natl. Inst. Adv. Industrial Sci. Technol. (AIST), Central#2,1-1-1, Umezono, Tsukuba, Ibaraki, 305-8568 Japan, a-sasou@aist.go.jp)

This paper describes a novel interface that enables the control of a wheelchair by disabled people without having to use their hands. Thus far, a voice-driven wheelchair has been developed for disabled people who find it difficult to move their hands but are able to speak. Such wheelchairs usually adopt a headset microphone in order to reduce noise interference. However, a headset microphone is not practical because disabled people cannot wear the headset independently. Another problem faced is that the microphone shifts away from the mouth due to involuntary movements. In the proposed system, however, we use a microphone array that is mounted on a wheelchair, so that the disabled people do not need to wear the microphone. The system is also robust to noises and is able to recognize a face's direction by estimating the position of a sound source such as the mouth. For instance, when disabled people need to turn left, they face left and make a sound that is not restricted to voice commands. Thus, the proposed system does not require disabled people to utter clear voice commands. This feature enables the application of the system to more general cases.

3pSP50. Decomposition of the time reversal operator (DORT) for extended objects. Jean-Luc Robert (Philips Res. North America, 345 Scarborough Rd., Briarcliff Manor, NY 1510, jean-luc.robert@philips.com) and Mathias Fink (Universite Paris 7, 75005 Paris, France)

The decomposition of the time reversal operator, known by the French acronym DORT, is widely used to detect, locate, and focus on scatterers in various domains such as underwater acoustics, medical ultrasound, or nondestructive evaluation. In the case of point scatterers, the theory is well understood: the number of nonzero eigenvalues is equal to the number of scatterers, and the eigenvectors correspond to the scatterers' Green function. In the case of extended objects, however, the formalism is not as simple. We show that, in the Fresnel approximation, analytical solutions can be found, and that the solutions are functions called prolate spheroidal wavefunctions. These functions have been studied in optics and information theory as a basis of band-limited and time-limited signals. The theoretical solutions are compared to simulation results and previous experimental results Most importantly, we justify the intuition that for an extended object, the number of nonzero eigenvalues is proportional to the number of resolution cells in the object. The case of 3-D objects imaged by a 2-D array will also be described.

Session 3pUW

Underwater Acoustics: Sound Propagation: Measurements and Modeling II

Nobuyuki Endoh, Cochair

Kanagawa Univ., Dept. of Engineering, 3-27-1 Rokkakubashi, Yokohama 221-8686, Japan

Martin Siderius, Cochair

Heat Light & Sound Research Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130

Contributed Papers

1:00

3pUW1. Measurements of temporal coherence of sound transmissions through shallow water. T. C. Yang (Naval Res. Lab., Washington, DC 20375)

Temporal coherence of acoustic signals propagating through fluctuating oceans has received much attention in the literature as the rate of signal coherence loss impacts the sonar performance. This paper presents the measurement results of temporal coherence of acoustic signals propagating through shallow water using data from three experiments in three different parts of the world. The data are broadband signals covering three frequency bands: low frequencies below 1.2 kHz, mid frequencies between 2 and 5 kHz, and high frequencies between 18 and 22 kHz. The source-receiver range covers 3, 5, 10, and 42 km. Measurement issues and environmental impact on the temporal coherence are discussed. The signal coherence time, defined by the time that the signal becomes decorrelated, is deduced from the data. The signal coherence time varies with the source-receiver range and the signal frequency and is significantly influenced by the medium fluctuations. The data can be fit with a simple power law dependence on the range, frequency, and sound speed standard deviations. [This work is supported by the U.S. Office of Naval Research.]

1:15

3pUW2. Comparison of measured signal coherence time in shallow and deep water. T. C. Yang (Naval Res. Lab., Washington, DC 20375)

Temporal coherence of acoustic signals determines the processing of a sonar system for various applications. This paper studies the frequency and range dependence of the temporal coherence time of signal propagation in deep water versus that in shallow water. The signal coherence time has been measured in deep water since the 1970s, in terms of a parameter called the phase rate (the rate of phase change of the signal). Signal coherence time in shallow water has received little attention until only recently. We compare measurement data from more than a dozen of experiments covering a wide range of frequencies at different source-receiver ranges in deep and shallow water. We find that the signal coherence time in both shallow and deep water can be fitted with a universal equation, which exhibits a $\frac{3}{2}$ power frequency dependence and $\frac{1}{2}$ power range dependence. The signal coherence time in shallow water is two to five times longer than the signal coherence time in deep water at the same frequency and range. The signal coherence times are comparable between shallow and deep water if measured in terms of the range-to-depth ratio. [This work is supported by the U.S. Office of Naval Research.]

1:30

3pUW3. Spatial coherence of the acoustic field for propagation ranges 500−1000 m in an East China Sea channel. Peter H. Dahl and Jee Woong Choi (Appl. Phys. Lab., Univ. of Washington, Seattle)

Measurements of the vertical spatial coherence at ranges 500–1000 m and for transmission frequencies 2–20 kHz were made as part of the Asian Sea International Acoustics Experiment (ASIAEX) conducted in the East

China Sea (depth 105 m). In an earlier work [J. Acoust. Soc. Am. **115**, 589–599 (2004)], ASIAEX spatial coherence measurements corresponding to arrivals interacting once with the sea surface were modeled using the van Cittert–Zernike theorem in conjunction with model estimates of bistatic scattering scattering from the sea surface. The theorem relates the spatial coherence of an observed field to the distant source intensity distribution associated with this field. Here, the approach from the 2004 work is extended to study the full multiple path arrival involving bistatic scattering from the sea surface, and coherent reflection plus bistatic scattering from the seabed. Model and data comparisons are made for the vertical direction and predications are made for coherence in horizontal, and horizontal-transverse direction.[Work supported by the ONR.]

1:45

3pUW4. Statistics of the Parseval sum, match filter, and instantaneous single-frequency transmissions in an ocean waveguide. Mark Andrews (Northeastern Univ., 409 Dana Res. Ctr., Boston, MA 02115-5000), Tianrun Chen, Sunwoong Lee, Nicholas Makris (MIT, Cambridge, MA 02139), and Purnima Ratilal (Northeastern Univ., Boston, MA 02115-5000)

During the ONR Main Acoustic Clutter Experiment of 2003 on the New Jersey Continental Shelf, short duration broadband pulses with 50- to 150-Hz bandwidths were transmitted at 50-s intervals from a moored source array in the 300- to 1500-Hz range. The signals were measured by a towed horizontal receiving array at varying ranges from the source. The broadband data are processed in a variety of ways including the use of Parseval sum and match-filter, and as an instantaneous single frequency transmission. The ping-to-ping fluctuations in the measured acoustic data are a result of changes in the waveguide modal intereference structure due to both motion of the array and presence of time-dependent random internal waves. Our analysis shows that the Parseval sum and match-filter outputs for the direct arrival have significantly smaller standard deviations compared to the instantaneous single frequency transmission. We account for these observations by modeling the broadband acoustic field intensity in both a static waveguide and with Monte Carlo simulations in a fluctuating waveguide, as well as with the use of statistical distributions for the broadband acoustic field. We present an approach for empirically quantifying the range-dependent degradation in the match-filter output in a multimodal waveguide.

2:00

3pUW5. Estimation for velocity dispersion of normal modes in a laboratory waveguide. Cao Zhengliang, Wang Suping, and Zhou Shihong (Hangzhou Appl. Acoust. Res. Inst., Hangzhou 310012 China)

As one character of normal modes, velocity dispersion defined as velocity dependent on frequency is related to the environment of a fluid and seabed in a shallow water waveguide. It is important to inverse geoacoustic parameters or analyze normal mode propagation in the waveguide. With waveforms measured by a line array in a laboratory waveguide, velocity dispersion of normal modes is estimated by Prony's

method. Because of the influence of the sum of many modes in a frequency, the estimated dispersions are disordered and discontinued. While a number of modes fewer than that, in fact, are previously selected, the results can be improved on velocity dispersion curves. On the other hand, clear curves can be obtained for the components of waveforms intercepted by an adaptive time window. These results are consistent with that of synthetic waveforms simulated by a normal mode program. Comparing the dispersions estimated from data in different depths, there are abundant high-order modes in this case. [Work supported by SKLF.]

2:15

3pUW6. PlaneRay: An underwater acoustic propagation model using ray tracing and plane wave reflection coefficients. Jens M. Hovem and Hefeng Dong (Dept. of Electron. and Telecommunications, Norwegian Univ. of Sci. and Technol., No. 7491, Trondheim, Norway)

Recently there has been a renewed interest in using ray tracing techniques for propagation modeling in underwater acoustics and this paper describes a newly developed model. A special and essential feature of this model is a unique sorting and interpolation routine for efficient determination of a large number of eigenrays and also for range-dependent environments. No rays are traced into the bottom and the bottom interaction is modeled by plane wave reflection coefficients. In the current version of the model the bottom structure consists of a fluid sediment layer over a solid half-space. Important considerations are the accuracy of the ray model and, in particular, how accurate the plane wave reflection coefficients can represent the effects of a layered bottom. This problem is analyzed by comparing the time and frequency domains solution of the ray model with models based on the wave number integration technique. There are two conflicting requirements since ray tracing is valid for high frequencies and the used plane wave reflection coefficient is valid for low frequencies where the sediment layers are thin compared with the acoustic wavelength. A number of illustrative examples are presented.

2:30

3pUW7. Ray-based acoustic localization of cavitation in a highly reverberant environment. Natasha A. Chang and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, 1231 Beal Rd., 2019 Lay Autolab, Ann Arbor, MI 48109)

Acoustic localization techniques are applied in a water tunnel to study low event rate cavitation in a vortical flow. The measured center frequency and bandwidth of the acoustic pulse from the collapse of a small isolated cavitation bubble are 10 kHz and 200 kHz, respectively, and the measured pulse duration is ~ 15–20 μs. Its timing and exact waveform is not known. An array of 16 receiving hydrophones is used to estimate the source location in the test section of the water tunnel. The direct path signal used for this effort can be buried in background hydrodynamic noise from the water tunnel. Fortunately some of the direct-path signal arrival is distinct enough for acoustic localization. The signal coda is not used in the present effort. The received direct-path signals are cross-correlated between hydrophones to obtain arrival time differences. These arrival time differences are used in conjunction with a simple ray-based acoustic model to estimate the source location in three dimensions. Once the cavitation bubble is located, the signal can be back propagated and part of the free space signal estimated and correlated with high-speed video of cavitation bubble dynamics. [Work sponsored by ONR codes 333 and 321.]

2:45

3pUW8. Bistatic invariance for active sonar systems. Lisa Zurk, Jorge Quijano, Manish Velankar (Portland State Univ., P.O. Box 751-ECE, Portland, OR 97207, zurkl@cecs.pdx.edu), and Dan Rouseff (Univ. of Washington, Seattle, WA 98105)

The concept of an invariant time-frequency relationship for passive acoustic intensity has been established through the consideration of interfering normal modes traveling in a shallow water channel. This formulation has received considerable attention in the past and has been applied to the interpretation of passive lofargrams. For bistatic active geometries, the acoustic energy propagates along two distinct pathways: source-to-target and target-to-receiver. In recent work we proposed that this more complicated propagation may also exhibit an invariant structure, thus suggesting potential for improved target tracking through the application of invariant-based track constraints. In this work, further exploration of this phenomenon is presented, with specific attention to the impact of the target scattering matrix that couples energy at the target between incoming and outgoing modes. Simulation results are presented for a variety of underwater environments to examine the effect on the invariant structure. Results are also presented and discussed from spectrograms produced from active sonar data obtained in shallow water environments.

3:00–3:15   Break

3:15

3pUW9. Time-frequency analysis of mode coupling in shallow-water sound propagation. Mohsen Badiey (Ocean Acoust. Lab., Univ. of Delaware, Newark, DE 19716), Valery Grigorev, Boris Katsnelson (Voronezh Univ., Voronezh 394006 Russia), and James Lynch (WHOI, Woods Hole, MA 02543)

Time-frequency (TF) analysis for propagation of broadband sound pulses in shallow water in presence of perturbations of waveguides parameters, such as internal waves, temperature fronts, or bottom variability is considered. Due to waveguide dispersion one can construct a TF diagram for individual normal modes using Gabor transform for each received pulse. If we have a temporal sequence of propagating pulses then we can find a temporal sequence of TF diagrams. It is shown that in the presence of spatial and temporal waveguide perturbations (such as the sound-speed profile in the water column or the sea bottom) structure of a TF diagram and its temporal behavior depends on waveguide parameters, on the source-receiver positions, and on the character of mechanisms influencing the sound propagation, such as mode coupling, horizontal refraction, or adiabatic regime. Dependence of the TF diagram on waveguide variability in geotime provides information about temporal behavior of the perturbations. Structure of TF diagram is analyzed for a shallow-water waveguide model as well as for the SWARM95 experimental data. [Work supported by ONR and RFBR.]

3:30

3pUW10. Propagation of impulsive broadband signals in a coastal ocean setting during the 1996 Shelfbreak Primer experiment. Steven E. Crocker (Naval Undersea Warfare Ctr., Newport, RI 02841), James H. Miller, Gopu R. Potty, Georges A. Dossot (Univ. of Rhode Island, Narragansett, RI 02882), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The 1996 Shelfbreak PRIMER experiment provided an opportunity to study the impulse response of a coastal ocean waveguide using broadband acoustic signals from explosive sources. Acoustic transmission paths in the vicinity of the shelfbreak included various orientations ranging from upslope to along-slope, including interactions with a shelfbreak front. Acoustic receptions were acquired with a 16-element vertical line array (VLA). In addition, thermistor string data acquired at the receiving array suggest that a group of high-amplitude, internal waves traversed the test site during this portion of the experiment. Modal decomposition of broadband signal arrivals was performed to investigate variation of the impulse response during source deployments. Variations in the received spectra of individual normal modes were compared with numerical predictions for impulsive broadband signals propagating in range-dependent environments constructed using oceanographic data available for the test site. The potential for mode coupling and three-dimensional effects to shape the modal and spectral distribution of received energy is being investigated. Numerical and experimental results will be discussed. [Work supported by Office of Naval Research.]

3p THU. PM

3pUW11. Underwater propagation at 16 Hz from French Polynesia to New Zealand: Echoes from islands, banks, and seamounts, and attenuation by seamounts. Marshall V. Hall (9 Moya Crescent, Kingsgrove, New South Wales 2208, Australia, marshallhall@optushome.com.au)

Underwater signals from two nuclear shots fired in Fangataufa and three in Mururoa during 1995/96 were recorded off northern New Zealand, a geoid distance of 4660 km. The signals have been bandpass filtered at 16 Hz and found to contain up to six echoes. The causative scattering seafloor features were estimated from the travel times. For some echoes there was ambiguity in the feature and hence some uncertainty in the deflection angle (DA). Amongst the six echoes, the DA ranged between 20 and 56 deg. In the two Fangataufa signals, echo intensities relative to the intensity of the direct pulse ranged between -44 and -25 dB, and this relative intensity (RI) decreased with increasing DA. The highest feature on the direct path from Fangataufa, a seamount whose peak is 1.6 km deep, appears to have had little effect on the signal. In the three Mururoa signals, there were fewer echoes, and RI did not vary significantly with DA. The highest feature on the direct path from Mururoa was a seamount whose peak is 0.7 km deep. This seamount attenuated the direct signal at New Zealand by approximately 23 dB, as well as paths to features at small angles from the direct.

4:00

3pUW12. Features of the broadband acoustic propagation in very shallow water. Seongwook Lee, Kyu-Chil Park, Jong Rak Yoon (Div. of Electron., Comput. and Telecommunication Eng., Pukyong Natl. Univ., 599-1 Daeyon-3dong, Namgu, Pusan 608-737, Korea), and Phil-Ho Lee (Agency for Defense Development, Jinhae, Gyungnam 645-600, Korea)

In shallow water, acoustic propagation is severely influenced by the property and structure of the bottom. We present the results of acoustic measurement conducted in a very shallow water of about 11-m depth to characterize the features of broadband acoustic propagation. In the measurements using both of the ship and projectors as sources, the signal was received at four hydrophones, two of them were located in the water and another two were located in the sediment. Broadband propagation losses derived from ship noise revealed the frequency band that has high losses around 120–140 Hz. Analysis results on the possible causes of high losses at that frequency band are presented.

4:15

3pUW13. A stochastic model for underwater acoustic channel impulse function in the Kauai experiment. Aijun Song, Jing Luo, and Mohsen Badiey (College of Marine Studies, Univ. of Delaware, Newark, DE 19716)

Modeling of underwater acoustic channel impulse function has been of interest to the research community. As is well known, the underwater acoustic wave propagation depends on a number of environmental parameters, such as the sound speed profile along the propagation track, sea surface condition, and source/receiver geometry. The fluctuations of these environmental parameters cause the variability of the acoustic wave propagation in the shallow water region. In this paper, a stochastic model for the channel impulse function is proposed. The characteristics of the each eigenray, such as arrival time, average intensity, and coherence time, are extracted from the Kauai experiment (2003). The correlation between the model parameters and the environmental variability will also be shown.

3pUW14. Measurements of 3-D propagation effects at a shelfbreak front. Kristy A. Moore, James H. Miller, Gopu R. Potty, Georges A. Dossot (Dept of Ocean Eng., Univ. of Rhode Island, 217 Sheet Bldg., Narragansett, RI 02882), Scott M. Glenn ((Rutgers Univ., NB, NJ 08901), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Acoustic data collected on two Webb Slocum gliders deployed by Rutgers University during the Shallow Water Experiment (SW06) on the continental shelf off New Jersey are presented. Acoustic recording systems were attached to the two gliders, which also collected oceanographic data. These gliders periodically surfaced for GPS fixes and data transfer via satellite phone. A major goal of these measurements was to quantify 3-D propagation effects including the horizontal analog to the classical Lloyd's mirror. The Lloyd's mirror effect is produced by the fact that fronts can totally internally reflect sound incident upon them at low grazing angles. The direct and reflected arrivals can add up to produce an interference pattern in range, potentially increasing the received intensity level by 6 dB [Lynch *et al.*, IEEE J. Ocean. Eng., in press]. The gliders crossed the shelfbreak front at ranges of approximately 20 km from the sources and in water depths of approximately 80 m. Measurements of the acoustic arrivals from 224 and 400 Hz tomography sources are compared to 3D modeling results. The increase in the intensity level predicted by the horizontal ray-vertical mode model is discussed. [Work supported by Office of Naval Research.]

4:45

3pUW15. Arrival time change of long distance sound travel in the central Pacific. Hanako Ogasawara, Toshiaki Nakamura (Dept. of Earth and Ocean Sci., Natl. Defense Acad., 1-10-20 Hashirimizu, Yokosuka, Kanagawa 239-8686, Japan), Hiroyuki Hachiya (Chiba Univ., Inage-ku, Chiba 269-8522, Japan), Hidetoshi Fujimori (Japan Agency for Marine-Earth Sci. and Technol., Yokosuka, Kanagawa 237-0061, Japan), and Koichi Mizutani (Univ. of Tsukuba, Tsukuba, Ibaraki 205-8573, Japan)

Ocean acoustic tomography (OAT) is a valuable method for observations of oceanographic phenomena over wide areas. A tomography experiment using seven 200-Hz transceivers at depths around 1000 m was executed in the central equatorial Pacific Ocean during January through December 2000 to monitor ocean phenomena related to El Niño and southern oscillation (ENSO). Because sound that propagates in the ocean takes various paths depending on its incident angle, the sound arrival times showed time lags. Arrival times of the first arrival signals were used to evaluate fluctuations throughout the year. Positional data of transceivers calculated from transponders indicated that major fluctuations of arrival times resulted from location movements of transceivers. The arrival times were corrected by calculating the distance using the positional data to remove distance effects in each transmission. Tidal effects were observed from spectral analyses of the corrected arrival times throughout the experimental period. Reciprocal arrival time differences between one pair of transceivers showed different changes during mid-September through November in comparison to other reciprocal arrival times of several paths. This change might have resulted from some kind of oceanic variation. Acoustic observations are advantageous for detecting various changes over long periods.

Session 4aAA

Architectural Acoustics: Measurement of Room Acoustics I

Fumiaki Satoh, Cochair

Chiba Inst. of Technology, Tsudanuma 2-17-1 Narashino-shi, Chiba 275-0016, Japan

Boaz Rafaely, Cochair

Ben Gurion Univ., Electrical and Computer Engineering Dept., 84105, Beer Sheva, Israel

Chair's Introduction—7:30

Invited Papers

7:35

4aAA1. Warped-time-stretched pulse: An acoustic test signal robust against ambient noise. Masanori Morise, Toshio Irino, Hideki Banno, and Hideki Kawahara (Wakayama Univ., 930, Sakaedani, Wakayama, 640-8510, Japan, s055068@sys.wakayama-u.ac.jp)

A new acoustic measurement signal that is a hybrid signal of time-stretched pulse (TSP), or lin-TSP, and logarithmic TSP (log-TSP) is proposed. The signal, referred to as warped-TSP [Morise *et al.*, IEICE Trans. Fundamentals, A, **J89-A**(1), 7–14 (2006)], has a single parameter to adjust for better measurements in accordance with ambient noise conditions. It also provides a means to eliminate harmonic distortions produced mainly by loudspeaker systems. In this lecture, the definition and features of the warped-TSP in comparison with the lin-TSP and log-TSP are introduced. The following were shown: (1) the relationship between the parameters, the amplitude frequency characteristics, and the effect on the harmonic distortion components; (2) a method to select the optimal parameters of the warped-TSP for a specific measuring environment; and (3) the experimental results for a series of impulse response measurements under different ambient noise conditions. Those results show that the proposed method outperformed the lin-TSP and log-TSP under all conditions in terms of SNR of the measured impulse response. [This research was supported partly by grants-in-aid for scientific research (15300061 and 15650032) and a grant from the Faculty of Systems Engineering at Wakayama University.]

7:55

4aAA2. Simultaneous estimation of reverberation times and their uncertainties from room impulse responses using a single-measurement procedure. Ning Xiang and Tomislav Jasa (Grad. Program in Architecture Acoust., and Dept. of Elec., Comput., and Systems Eng., Rensselaer Polyt. Inst, Troy, NY 12180)

Accurate measurements of reverberation times are of fundamental importance in room acoustics. A number of test procedures for characterizing acoustics in performing arts venues, quantifying acoustic properties of materials in chamber measurements, rely on experimental determination of reverberation times. In addition, decay-time estimation in acoustically coupled spaces has been found to be very demanding. Our recent work has demonstrated that model-based Bayesian approaches [Xiang *et al.*, J. Acoust. Soc. Am. **110**, 1415–1424 (2001); **113**, 2685–2697 (2003); **117**, 3705–3715 (2005)] can be very useful for such analysis in architectural acoustics measurements. This paper discusses the recent development of probabilistic tools for estimating both reverberation (decay) times and their uncertainties within Bayesian framework. This work shows that Bayesian probabilistic inference can be used as a useful tool for sound energy decay analysis in both single-space halls and coupled spaces. Bayesian decay analysis simultaneously provides architectural acousticians with reverberation times, diverse decay times, related derivations, and interdependencies to quantify uncertainties of the estimation from a single measurement of room impulse responses followed by Schroeder backward integrations.

8:15

4aAA3. Permissible number of synchronous averaging times to obtain reverberation time from impulse response under time-variance conditions. Fumiaki Satoh, Yukiteru Hayashi (Chiba Inst. of Technol., Tsudanuma 2-17-1, Narashino-chi, Chiba, 275-0016, Japan), Shinichi Sakamoto (Univ. of Tokyo, Meguro-ku, Tokyo, 153-8505, Japan), and Hideki Tachibana (Chiba Inst. of Technol., Narashino-shi, Chiba, 275-0016, Japan)

In the measurement of room impulse response, the synchronous averaging technique and such new methods as the MLS and the swept-sine methods are being widely used to improve the signal-to-noise ratio. In actual measurement conditions, however, the air in a room is continuously moving and the temperature is changing to some degree. The measured value of the reverberation time in such a room tends to be shorter at higher frequencies when applying the synchronous averaging. Therefore, the assumption of a time invariant has to be carefully considered, and, on this point, some research has been conducted to date. We also have reported various research results concerning the impulse response measurement under the time-variance conditions. In this paper, the permissible number of synchronous averaging times for reverberation measurement is studied through some field experiments. In each field, many

3223 J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ **3223**

time impulse response measurements were taken between a fixed pair of sound source and receiving positions by the swept-sine method, without averaging. After the measurements, the characteristics and the extent of the time-variance under measuring were estimated by a short-term running cross-correlation function between each impulse response. The influence of the time variance on the synchronous averaging result was studied based on the estimated time variance.

8:35

4aAA4. Selection of receiving positions suitable for evaluating acoustical parameters. Taeko Akama, Hisaharu Suzuki, and Akira Omoto (Omoto Lab., Dept. of Acoust. Design, Faculty of Design, Kyushu Univ., Shiobaru 4-9-1, Minami, Fukuoka 811-8540, Japan)

Many physical parameters show characteristics of large sound fields such as concert halls. Some of them are adopted in the Annex of ISO 3382. That definition is clearly provided in ISO. However, practical measurement methods for them remain obscure. Our research is intended to examine an effective selection method of receiving positions based on the distribution of acoustical parameters in a real field. For that purpose, impulse responses are measured at more than 1400 seat positions to elucidate the distribution of acoustical parameters in an existing concert hall. The acoustical parameters, which are reverberation time, early decay time, clarity, and center time at each seat, are then calculated for 500-Hz, 1-kHz, and 2-kHz octave bands. The distributions of reverberation time are quite even at all seats. However, the distributions of other parameters show symmetrical patterns at 500 Hz. At 1 and 2 kHz frequencies, the distributions show asymmetrical patterns in this hall. Based on the results obtained in this study, an effective method to select the receiving position can be proposed.

8:55

4aAA5. Hybrid measurement method in room acoustics using dodecahedron speakers and a subwoofer. Hideo Miyazaki (Ctr. for Adv. Sound Technologies, Yamaha Corp., 203 Matsunokijima, Iwata, Shizuoka 438-0192, miyazaki@beat.yamaha.co.jp)

A dodecahedron speaker is usually utilized for measurement in room acoustics under the hypothesis of omni directional point source. But generally speakers used for a dodecahedron speaker cannot playback low-frequency sound such as under 100 Hz, which is important especially for auralization, while the one constructed of units with large diameter to support low-frequency sounds cannot be considered as an omni-directional speaker in high frequencies. To meet these requirements, a hybrid system combining a dodeca-hedron speaker and a subwoofer has been developed and actually used for measurements of impulse responses in acoustical design of concert halls. The summary of this method will be presented. The feasibility of this method will be also discussed while evaluating the measurement results in concert halls by changing measurement conditions such as speaker locations and comparing these results with those of conventional methods.

9:15

4aAA6. The perception of apparent source width and its dependence on frequency and loudness. Ingo B. Witew and Johannes A. Buechler (Inst. of Tech. Acoust., RWTH Aachen Univ., Templergraben 55, 52066 Aachen, Germany)

While it is widely accepted that apparent source width (ASW) is an important factor in characterizing the acoustics of a concert hall, there is still a lively discussion on how to refine the physical measures for ASW. A lot of experience has been gathered with interaural-cross-correlation and lateral-sound-incidence measures during the last years. As a result it was learned that different frequencies contribute differently to the perception of ASW and that the level of a sound also influences the perception of the apparent width of a source. With many technical measures having an influence on the perceptual aspect of ASW, the design of psychometric experiments becomes challenging as it is desirable to avoid the interaction of different objective parameters. In the experiments for the study presented, the perception of ASW is investigated for frequencies ranging from 100 Hz to 12.5 kHz at different levels of loudness. It is shown how the frequency and the loudness of a sound influence the perception of ASW.

9:35

4aAA7. Sound source with adjustable directivity. Gottfried K. Behler (Inst. fuer Technische Akustik, RWTH Aachen Univ., D-52056 Aachen, Germany)

Omni-directional sound sources are used to measure room-acoustical parameters in accordance with ISO 3382. To record a detailed room impulse response (RIR) with the aim of auralization, an extended frequency range is required that is not covered by the often-used building acoustics sound sources. To obtain this target, a loudspeaker with dedicated sources for low, mid, and high frequencies was designed, providing a smooth omni-directionality up to 6 kHz and a usable frequency range from 40 Hz up to 20 kHz. However, a realistic auralization of sources like musical instruments is not possible with an omni-directional measured RIR. To include the directional characteristics of instruments in the measuring setup, the directivity of the sound source has to be frequency dependent and must be matched to the (measured) directivity of the real instrument. This can be obtained by using a dodecahedron loudspeaker with independently operating systems and an appropriate complex FIR filtering of the frequency response of each driver. The directivity is a result of parameters like magnitude and phase and the interference sum of all systems. To create the appropriate directivity, optimization algorithms are used to achieve minimum error between measured and adapted directivity.

10:15

4aAA8. Objective measures for evaluating tonal balance of sound fields.   Daiji Takahashi   (Dept. of Urban and Environ. Eng., Kyoto Univ., Kyoto Univ. Katsura, Nishikyo-ku, Kyoto 615-8540, Japan, tkhs@archi.kyoto-u.ac.jp), Kanta Togawa   (FUJITEC Co., Ltd., Hikone Shiga 522-8588, Japan), and Tetsuo Hotta   (YAMAHA Corp., Hamamatsu Shizuoka 430-8650, Japan)

The purpose of this study is to derive objective measures, which can well represent the characteristics of the sound field regarding the tonal balance corresponding to our hearing sense. Two kinds of listening test were conducted in the form of paired comparison, in which subjects were tested using sound fields produced by convoluting some anechoic music sources with some impulse responses. In the first listening test, impulse responses were calculated theoretically for a simple structure of sound field having a direct sound and reflections, and, in the second test, impulse responses were measured at the various seats of existing concert halls. In the latter case, impulse responses which give almost the same reverberance were used for the listening tests. From this investigation, it is found that one objective measure called the DL (deviation of level) has a possibility of an effective measure, which can be used as an appropriate measure for evaluating the tonal balance of sound fields. The index DL is calculated from the data based on the logarithmic scale in both the frequency and magnitude. This fact is not inconsistent with the past findings that human response corresponds to a logarithmic scale of stimulus.

10:35

4aAA9. Measuring impulse responses containing complete spatial information.   Angelo Farina, Paolo Martignon, Andrea Capra, and Simone Fontana   (Industrial Eng. Dept., Univ. of Parma, via delle Scienze 181/A, 43100 Parma, Italy)

Traditional impulse response measurements did capture limited spatial information. Often just omnidirectional sources and microphones are employed. In some cases it was attempted to get more spatial information employing directive transdudcers: known examples are binaural microphones, figure-of-8 microphones, and directive loudspeakers. However, these approaches are not scientifically based and do not provide an easy way to process and visualize the spatial information. On the other side, psychoacoustics studies demonstrated that "spatial hearing" is one of the dominant factors for the acoustic quality of rooms, particularly for theatres and concert halls. Of consequence, it is necessarily to reformulate the problem entirely, describing the transfer function between a source and a receiver as a time/space filter. This requires us to "sample" the impulse response not only in time, but also in space. This is possible employing spherical harmonics for describing, with a predefined accuracy, the directivity pattern of both source and receiver. It is possible to build arrays of microphones and of loudspeakers, which, by means of digital filters, can provide the required directive patterns. It can be shown how this makes it possible to extract useful information about the acoustical behavior of the room and to make high-quality auralization.

10:55

4aAA10. Spherical and hemispherical microphone arrays for capture and analysis of sound fields.   Ramani Duraiswami, Zhiyun Li, Dmitry N. Zotkin, and Elena Grassi   (Perceptual Interfaces and Reality Lab., Inst. for Adv. Comput. Studies, Univ. of Maryland, College Park, MD 20742)

The capture of the spatial structure of a sound field and analysis is important in many fields including creating virtual environments, source localization and detection, noise suppression, and beamforming. Spherical microphone arrays are a promising development to help achieve such capture and analysis, and have been studied by several groups. We develop a practical spherical microphone array and demonstrate its utility in applications for sound capture, room measurement and for beamforming and tracking. To accommodate equipment failure and manufacturing imprecision we extend their theory to handle arbitrary microphone placement. To handle speech capture and surveillance we describe the development of a new sensor, the hemispherical microphone array. For each array the practical performance follows that predicted by theory. Future applications and improvements are also discussed. [Work supported by NSF.]

11:15

4aAA11. High-order wave decomposition using a dual-radius spherical microphone array.   Boaz Rafaely, Ilya Balmages, and Limor Eger   (Dept. of Elec. and Comput. Eng., Ben-Gurion Univ. of the Negev, Beer-Sheva 84105, Israel)

The acoustic performance of an auditorium is influenced by the manner in which sound propagates from the stage into the seating areas. In particular, the spatial and temporal distribution of early reflections is considered important for sound perception in the auditorium. Previous studies presented measurement and analysis methods based on spherical microphone arrays and plane-wave decomposition that could provide information on the direction and time of arrival of early reflections. This paper presents recent results of room acoustics analysis based on a spherical microphone array, which employs high spherical harmonics order for improved spatial resolution, and a dual-radius spherical measurement array to avoid ill-conditioning at the null frequencies of the spherical Bessel function. Spatial-temporal analysis is performed to produce directional impulse responses, while time-windowed space-frequency analysis is employed to detect direction of arrival of individual reflections. Experimental results of sound-field analysis in a real auditorium will also be presented.

<div style="float: right">**4a FRI. AM**</div>

4aAA12. Impulse response measurement system and its recent applications. Kazuhiro Takashima, Hiroshi Nakagawa, Natsu Tanaka, and Daiki Sekito (1-21-10, Midori, Sumida-Ku, Tokyo 130-0021, Japan)

Our impulse response measurement system has been developed for ten years. During this decade, the environment related to this measurement has changed significantly. In this article, the features and notes on the measurement system using the sound card, and our brand new system, which is expanded for multichannel inputs, will be presented. Finally, a new technique, which combines multi-channel impulse response measurement and signal processing with microphone array, will be presented. The microphone array was designed for noise analysis for automobile interiors. The array consists of 31 microphones on the surface of an acoustically hard sphere. Moreover, 12 cameras are arranged on the surface of the sphere to take photos. Some applications and future development will be presented.

FRIDAY MORNING, 1 DECEMBER 2006   KOHALA/KONA ROOM, 8:00 TO 11:45 A.M.

Session 4aAB

Animal Bioacoustics: Marine Mammal Acoustics I

Paul E. Nachtigall, Chair
Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

Contributed Papers

8:00

4aAB1. Development of evoked-potential audiometry in odontocetes. Alexander Supin (Inst. of Ecology and Evolution, 33 Leninsky prospect, 119071 Moscow, Russia)

Evoked-potential methods are widely used for investigation of hearing in whales, dolphins, and porpoises. For this purpose, mostly the auditory brainstem response (ABR) or rhythmic trains of ABRs, the envelope-following response (EFR), are used. Although very productive, these methods require further elaboration. (*i*) Traditionally the EFR is provoked by sinusoidally amplitude-modulated tones (SAM). SAM stimuli have narrow frequency band, which makes them little effective to produce the EFR, because response amplitude depends on the stimulus bandwidth. A solution of the problem is the use of trains of short tone pips instead of SAM tones. Such stimuli produce several times higher EFR than SAM tones. This makes the threshold determination much more confident and precise. The effect is achievable at stimulus bandwidths, which still do not influence negatively the precision of attribution of the threshold to a certain frequency. (*ii*) To extract low-amplitude evoked potentials from noise, the average technique is traditionally used. This operation returns a mean value of averaged traces. Effectively diminishing stationary noise, this method poorly eliminates big artifacts, which may spoil the record even it if appeared once or twice during acquisition. With this respect, computation of the median instead of mean is much more effective.

8:15

4aAB2. Towards a predictive model of noise-induced temporary threshold shift for an amphibious marine mammal, the California sea lion (*Zalophus californianus*). David Kastak, Marla M. Holt, Jason Mulsow, Colleen J. Reichmuth Kastak, Ronald J. Schusterman (UCSC Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060), and Brandon L. Southall (Natl. Marine Fisheries Service, Silver Spring, MD 20910)

A California sea lion that had previously been tested under water was assessed for noise-induced temporary threshold shift (TTS) in air. One hundred ninety-two controlled exposures of octave-band noise centered at 2.5 kHz were conducted over a 3-year period. The noise was varied in level (to 133 dB SPL *re*: 20 μPa) and duration (to 50 min) to generate a variety of equal sound exposure levels (SELs). Behavioral psychophysics was used to measure hearing sensitivity at 2.5 kHz before, immediately following, and 24 h following noise exposure. The levels of threshold shifts obtained ranged up to 30 dB. In cases where TTS exceeded 20 dB, thresholds were obtained at regular intervals until recovery occurred. The average slope of the long-term recovery function was 10 dB per log-(minute). Results show that the threshold shifts correlated with SEL; however, the equal-energy trading rule did not apply in all circumstances, with exposure duration contributing more than exposure level. Repeated testing showed no evidence of a permanent threshold shift at 2.5 kHz or octave higher. The amphibious sea lions appear to be equally susceptible to noise in air and under water, provided that the exposure levels are referenced to the subjects thresholds in both media.

8:30

4aAB3. Electrophysiological investigation of temporal resolution in three pinniped species: Adaptive implications. Jason Mulsow and Colleen Reichmuth Kastak (Univ. of California Santa Cruz, Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060)

Electrophysiological studies of auditory temporal processing in marine mammals have traditionally focused on the role of highly refined temporal resolution in dolphin echolocation. Studies in manatees, however, have found their temporal resolution to be better than expected, leading to speculation that such capabilities are an adaptation for underwater sound localization. This study measured the ability of auditory brainstem responses to follow rhythmic click stimuli in California sea lions (*Zalophus californianus*), harbor seals (*Phoca vitulina*), and northern elephant seals (*Mirounga angustirostris*). Trains of 640-s clicks were presented in air at repetition rates of 125–1500 per second and averaged rate-following responses were recorded. Rate-following responses were detected in both the harbor seal and the sea lion at rates up to 1000 clicks per second, indicating that pinnipeds, like manatees, possess temporal resolution greater than humans but inferior to dolphins. While this finding might support an underwater sound localization hypothesis, comparable results were obtained in preliminary testing of a dog (*Canis familiaris*), suggesting that increased temporal resolution in pinnipeds may not be the result of the evolutionary pressure of an aquatic environment, but rather a result of increased high-frequency hearing essential to mammalian sound localization. [Work supported by NOPP, ONR, and NMFS.]

4aAB4. Click and tone-pip auditory evoked potentials in a large marine mammal, the northern elephant seal. Dorian S. Houser (BIOMIMETICA, 7951 Shantung Dr., Santee, CA 92071) and James J. Finneran (Space and Naval Warfare Systems Ctr., San Diego, CA 92152)

The use of auditory-evoked potentials (AEPs) to study the hearing of mysticete whales is challenged by access to animals, their large size, and proportionately smaller brain relative to odontocetes. One means by which AEP techniques can be adapted to these larger animals is by application to more readily available proxy species. The northern elephant seal (*Mirounga angustirostris*) is a large pinniped, potentially in excess of 2000 kg, with a thick dermis, large skull, relatively small auditory nerve, and a low-frequency vocal communication system. AEP collection in elephant seals provides similar challenges to those of the mysticetes but at a scale that provides a greater opportunity for success. AEP tests were conducted on northern elephant seals at Año Nuevo State Reserve, the natural haul-out site of the elephant seal. Subjects were chemically immobilized with tiletamine/zolazepam and chemical restraint was maintained with bolus injections of ketamine. Click-evoked potentials were collected from four weanling and two adult male elephant seals and tone-pip-evoked potentials were collected from a 2-year-old female. Results demonstrate that AEPs can be recorded from large pinniped species, providing a step towards the application of similar techniques to larger cetacean species.

4aAB5. Acoustic field measurements and bottlenose dolphin hearing thresholds using single-frequency and frequency-modulated tones. James J. Finneran (U.S. Navy Marine Mammal Program, SPAWARSYSCEN San Diego, Code 2351, 49620 Beluga Rd., San Diego, CA 92152, james.finneran@navy.mil) and Carolyn E. Schlundt (EDO Professional Services, San Diego, CA 92110)

Studies of underwater hearing are often hampered by the behavior of sound waves in small experimental tanks. At lower frequencies, tank dimensions are often not sufficient for free-field conditions, resulting in large spatial variations of sound pressure. These effects may be mitigated somewhat by increasing the frequency bandwidth of the sound stimulus, so effects of multipath interference average out over many frequencies. In this study, acoustic fields and bottlenose dolphin (*Tursiops truncatus*) hearing thresholds were compared for pure-tone and frequency-modulated stimuli. Experiments were conducted in a vinyl-walled, seawater-filled pool approximately $4 \times 5 \times 1.5$ m. Sound stimuli consisted of 500-ms tones at 13 carrier frequencies between 1 and 100 kHz. Frequency-modulated stimuli featured both linear and sinusoidal modulating waveforms with 5%, 10%, and 20% bandwidths. Acoustic fields were measured (without the dolphin present) at three depths over a 60×65-cm grid with a 5-cm spacing. Hearing thresholds were measured using a behavioral response paradigm and up/down staircase technique. Frequency-modulated stimuli with a 10% bandwidth resulted in significant improvements to the sound field without substantially affecting the dolphins hearing thresholds. [Work supported by ONR.]

4aAB6. Hearing frequency selectivity in four species of toothed whales as revealed by the evoked-potential method. Vladimir Popov (Inst. of Ecology and Evolution, 33 Leninsky Prosp., 119071 Moscow, Russia popov_vl@sevin.ru)

Frequency tuning curves were obtained using a tone-tone simultaneous masking paradigm in conjunction with the evoked potential recording. The masker was a continuous tone and the test was a sinusoidal amplitude-modulated (SAM) tonal signal, which evoked the envelope following response (EFR). The EFR was recorded in unanaesthetized animals from a head surface with the use of suction-cup electrodes. The obtained tuning curves featured very sharp tuning with Q(ERB) (quality estimated by the equivalent rectangular bandwidth) from 35 in *Tursiops truncatus* to nearly 50 in *Delphinapterus leucas*. This acuteness is several times better than in humans and many animals. The Q(ERB) dependence on probe frequency could be approximated by regression lines with a slope from 0.18 in *Tur-*

siops trucatus to 0.83–0.86 in *Phocoena phocoena* and *Neophocoena phocoenoides*. Thus, the frequency representation in the odontocete auditory system may be either near constant quality (in Tursiops) or near constant bandwidth (in porpoises). [Work supported by The Russian Foundation for Basic Research and Russian President Grant.]

4aAB7. Growth and recovery of temporary threshold shifts in a dolphin exposed to midfrequency tones with durations up to 128 s. Carolyn E. Schlundt (EDO Professional Services, 3276 Rosecrans St., San Diego, CA 92110, carolyn.melka@edocorp.com), Randall L. Dear (Sci. Applications Intl. Corp., San Diego, CA 92110), Donald A. Carder, and James J. Finneran (Space and Naval Warfare Systems Ctr., San Diego, San Diego, CA 92152)

Auditory thresholds at 4.5 kHz were measured in a bottlenose dolphin (*Tursiops truncatus*) before and after exposure to midfrequency tones at 3 kHz. Experiments were conducted in relatively quiet pools with low ambient noise levels at frequencies above 1 kHz. Behavioral hearing tests allowed for thresholds to be routinely measured within 4 min postexposure, and tracked recovery for at least 30 min postexposure. Exposure durations ranged from 4 to 128 s at sound pressure levels ranging from 149 to 200 dB *re*: 1 μPa. Sound exposure levels ranged from 155 to 217 dB *re*: 1 μPa^2/s. Temporary threshold shifts at 4 min postexposure (TTS$_4$) of up to 23 dB were observed. All thresholds recovered to baseline and pre-exposure levels, most within 30 min of exposure. [Work supported by the U.S. ONR.]

4aAB8. Auditory brainstem response recovery rates during double-pulse presentation in the false killer whale (*Pseudorca crassidens*): A mechanism of automatic gain control? Paul E. Nachtigall (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Alexander Ya. Supin (Russian Acad. of Sci., Moscow, Russia), and Marlee Breese (Hawaii Inst. of Marine Biol., Kailua, HI 96734)

The outgoing echolocation pulse and the return echo response can be approximately examined in the auditory system of an echolocating animal by presenting two pulses and determining the forward-masking effect of the first pulse on the response to the second pulse using auditory-evoked potential procedures. False killer whale, *Pseudorca crassidens*, auditory brainstem responses (ABR) were recorded using a double-click stimulation paradigm specifically measuring the recovery of the second (test) response (to the second click) as a function of the length of the interclick interval (ICI) following various levels of the first (conditioning) click. At all click intensities, the slopes of the recovery functions were almost constant: 0.60.8 V per ICI decade. Therefore, even when the conditioning-to-click level ratio was kept constant, the duration of recovery was intensity dependent: the higher intensity the longer the recovery. The conditioning-to-test-click level ratio strongly influenced the recovery time: the higher the ratio, the longer the recovery. This dependence was nearly linear, using a logarithmic ICI scale with a rate of 2530 dB per ICI decade. These data were used for modeling the interaction between the emitted click and the echo in the auditory system during echolocation.

4aAB9. Temporary threshold shifts in the bottlenose dolphin (*Tursiops truncatus*), varying noise duration and intensity. T. Aran Mooney (Dept. of Zoology and Hawaii Inst. of Marlne Biol., Univ. of Hawaii, 46-007 Lilipuna Rd., Kaneohe, HI 96744), Paul E. Nachtigall, Whitlow W. L. Au, Marlee Breese, and Stephanie Vlachos (Univ. of Hawaii, Kaneohe, HI 96744)

There is much concern regarding increasing noise levels in the ocean and how it may affect marine mammals. However, there is a little information regarding how sound affects marine mammals and no published

data examining the relationship between broadband noise intensity and exposure duration. This study explored the effects of octave-band noise on the hearing of a bottlenose dolphin by inducing temporary hearing threshold shifts (TTS). Sound pressure level (SPL) and exposure duration were varied to measure the effects of noise duration and intensity. Hearing thresholds were measured using auditory evoked potentials before and after sound exposure to track and map TTS and recovery. Shifts were frequency dependent and recovery time depended on shift and frequency, but full recovery was relatively rapid, usually within 20 min and always within 40 min. As exposure time was halved, TTS generally occurred with an increase in noise SPL. However, with shorter, louder noise, threshold shifts were not linear but rather shorter sounds required greater sound exposure levels to induce TTS, a contrast to some published literature. From the data a novel algorithm was written that predicts the physiological effects of anthropogenic noise if the intensity and duration of exposure are known.

10:30

4aAB10. Estimates of bio-sonar characteristics of a free-ranging Ganges river dolphin. Tamaki Ura, Harumi Sugimatsu, Tomoki Inoue (Underwater Technol. Res. Ctr., Inst. of Industrial Sci., Univ. of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan), Rajendar Bahl (IIT Delhi, New Delhi 110016, India), Junichi Kojima (KDDI R&D Labs. Inc., Saitama 356-8502, Japan), Tomonari Akamatsu (Fisheries Res. Agency, Ibaraki 314-0408, Japan), Sandeep Behera (Freshwater & Wetlands Prog., New Delhi 110003, India), Ajit Pattnaik, Muntaz Kahn (Chilika Dev. Authority, Orissa, India), Sudhakar Kar, Chandra Sekhar Kar (Off. of the Princip. CCF (Wildlife) & Chief Wildlife Warden, Blaubaneswar 751012, India), Tetsuo Fukuchi, Hideyuki Takashi (System Giken Co. Ltd., Kanagawa 253-0085, Japan), and Debabrata Swain (Simpilpal Biosphere and Tiger Reserve, Orissa, India)

This paper reports the first known studies of the bio-sonar characteristics of an isolated free-ranging Ganges river dolphin, *Platanista gangetica*. The animal preferred to roam in a deeper tract of the otherwise shallow river. The click sounds of the dolphin were recorded over a period of 2 days on a 3.2-meter-long high-frequency hydrophone array composed of three hydrophones forming an equispaced linear array and another two hydrophones in conjunction with the central hydrophone forming an SSBL triangular array in a plane perpendicular to the array axis. The array was deployed both in horizontal and vertical configurations. The array structure provided 3-D measurements of the source location through measurement of the interelement time delay. Bio-sonar characteristics such as click duration, bandwidth, and interclick intervals in click trains have been reported. Measurements of dolphin track and the relative click levels on the array hydrophones have been used to obtain a preliminary characterization of the animal's beam pattern.

10:45

4aAB11. Discriminating between the clicks produced by a bottlenose dolphin when searching for and identifying an object during a search task. Sandra Bohn, Stan Kuczaj (Univ. of Southern Mississippi, 118 College Dr., #5025, Hattiesburg, MS 39406, sandra.bohn@usm.edu), and Dorian Houser (BIOMIMETICA, Santee, CA 92071)

Clicks collected from an echolocating bottlenose dolphin completing a search task were compared in order to determine if the clicks produced when the dolphin was acquiring the target differed from the clicks produced when the dolphin was searching for the target. The clicks produced by a free-swimming dolphin completing the search task were recorded using a biosonar measurement tool (BMT), an instrumentation package carried by the dolphin that collected both the outgoing clicks and the returning echoes. A discriminant function analysis classified the clicks as search or acquisition using the variables of peak-to-peak amplitude, duration, peak frequency, center frequency, and bandwidth. The acquisition clicks were classified more accurately than the search clicks. Acquisition clicks and search clicks were significantly different across all five of the variables. These results suggest that the clicks produced by bottlenose dolphins acquiring a target are different than those produced by dolphins searching for a target.

11:00

4aAB12. Echo highlight amplitude and temporal difference resolutions of an echolocating *Tursiops truncatus*. Mark W. Muller, Whitlow W. L. Au, Paul E. Nachtigall, Marlee Breese (Marine Mammal Res. Program, Hawai'i Inst. of Marine Biol., 46-007 Lilipuna Rd., Kaneohe, HI 96744), and John S. Allen III (Univ. of Hawai'i at Manoa, Honolulu, HI 96822)

A dolphin's ability to discriminate targets may greatly depend on the relative amplitudes and the time separations of echo highlights within the received signal. Previous experiments with dolphins have varied the physical parameters of targets, but did not fully investigate how changes in these parameters corresponded with the composition of the scattered acoustic waveforms and the dolphin's subsequent response. A novel experiment utilizes a phantom echo system to test a dolphin's detection response of relative amplitude differences of secondary echo highlights and the time separation differences of all the echo highlights both within and outside the animal's integration window. By electronically manipulating these echoes, the underlying acoustic classification cues can be more efficiently investigated. In the first study, the animal successfully discriminated between a standard echo signal and one with the middle highlight amplitude at -7 dB. When the middle highlight amplitude was raised to -6 dB, the animal's discrimination performance radically dropped to 65%. This study suggests the animal may not be as sensitive to the secondary echo highlights as previously proposed. The experiments were repeated for the trailing highlight amplitude and the time separations between the primary and middle highlights and the middle and trailing highlights.

11:15

4aAB13. A background noise reduction technique for improving false killer whale (*Pseudorca crassidens*) localization. Craig R. McPherson, Owen P. Kenny, Phil Turner (Dept. of Elec. and Comput. Eng., James Cook Univ., Douglas 4811, Queensland, Australia), and Geoff R. McPherson (Queensland Dept. of Primary Industries and Fisheries, Cairns, 4870, Queensland Australia)

The passive localization of false killer whales (*Pseudorca crassidens*) in acoustic environments comprised of discontinuous ambient, anthropogenic, and animal sounds is a challenging problem. A background noise reduction technique is required to improve the quality of sampled recordings, which will assist localization using auditory modeling and signal correlation at extended ranges. The algorithm developed meets this requirement using a combination of adaptive percentile estimation, a median-based tracker, and Gaussian windowing. The results indicate successful improvement of the signal-to-noise ratio, and consequently a significant increase in the detection range of false killer whales in acoustically complex environments.

11:30

4aAB14. Analysis of Australian humpback whale song using information theory. Jennifer L. Miksis-Olds, John R. Buck (School for Marine Sci. and Technol., Univ. of Massachusetts Dartmouth, New Bedford, MA 02744, jmiksis@umassd.edu), Michael J. Noad (Univ. of Queensland, St. Lucia, QLD 4072 Australia), Douglas H. Cato (Defence Sci. & Tech. Org., Pyrmont, NSW 2009 Australia), and Dale Stokes (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

Songs produced by migrating whales were recorded off the coast of Queensland, Australia over 6 consecutive weeks in 2003. Approximately 50 songs were analyzed using information theory techniques. The average length of the songs estimated by correlation analysis was approximately 100 units, with song sessions lasting from 300 to over 3100 units. Song entropy, a measure of structural constraints and complexity, was estimated using three different methodologies: (1) the independently identically distributed model; (2) first-order Markov model; and (3) the nonparametric sliding window match length (SWML) method, as described in Suzuki *et al.* [J. Acoust. Soc. Am. **119**, 1849 (2006)]. The analysis finds the songs of migrating Australian whales are consistent with the hierarchical structure proposed by Payne and McVay (Science **173**, 585−597 (1971)], and

recently confirmed by Suzuki *et al.* for singers on the breeding grounds. Both the SWML entropy estimates and the song lengths for the Australian singers were lower than that reported by Suzuki *et al.* for Hawaiian whales in 1976−1978. These lower SWML entropy values indicate a higher level of predictability within songs. The average total information in the Australian sequence of song units was approximately 35 bits/song. Aberrant songs (10%) yielded entropies similar to the typical songs. [Sponsored by ONR and DSTO.]

FRIDAY MORNING, 1 DECEMBER 2006

KAHUKU ROOM, 7:55 A.M. TO 12:00 NOON

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration: Interaction of Cavitation Bubbles with Cells and Tissue

John S. Allen, Cochair

Univ. of Hawaii, Dept. of Mechanical Engineering, 2540 Dole St., Honolulu, HI 96822

Yoshiki Yamakoshi, Cochair

Gunma Univ., Faculty of Engineering, 1-5-1 Tenjin-cho, Kiryu-shi, Gunma 376-8515, Japan

Chair's Introduction—7:55

Invited Papers

8:00

4aBB1. Ultra-high-speed imaging of bubbles interacting with cells and tissue. Michel Versluis, Philippe Marmottant, Sascha Hilgenfeldt, Claus-Dieter Ohl (Phys. of Fluids, Univ. of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands), Chien T. Chin, Annemieke van Wamel, Nico de Jong (Erasmus MC, 3000 DR Rotterdam, The Netherlands), and Detlef Lohse (Univ. of Twente, 7500 AE Enschede, The Netherlands)

Ultrasound contrast microbubbles are exploited in molecular imaging, where bubbles are directed to target cells and where their high-scattering cross section to ultrasound allows for the detection of pathologies at a molecular level. In therapeutic applications vibrating bubbles close to cells may alter the permeability of cell membranes, and these systems are therefore highly interesting for drug and gene delivery applications using ultrasound. In a more extreme regime bubbles are driven through shock waves to sonoporate or kill cells through intense stresses or jets following inertial bubble collapse. Here, we elucidate some of the underlying mechanisms using the 25-Mfps camera Brandaris128, resolving the bubble dynamics and its interactions with cells. We quantify acoustic microstreaming around oscillating bubbles close to rigid walls and evaluate the shear stresses on nonadherent cells. In a study on the fluid dynamical interaction of cavitation bubbles with adherent cells, we find that the nonspherical collapse of bubbles is responsible for cell detachment. We also visualized the dynamics of vibrating microbubbles in contact with endothelial cells followed by fluorescent imaging of the transport of propidium iodide, used as a membrane integrity probe, into these cells showing a direct correlation between cell deformation and cell membrane permeability.

8:20

4aBB2. Sonoporation: Mechanisms of cell membrane perforation and rapid resealing. Nobuki Kudo and Katsuyuki Yamamoto (Grad. School of Information Sci. and Technol., Hokkaido Univ., Sapporo 060-0814 Japan, kudo@bme.ist.hokudai.ac.jp)

Sonoporation is a technique for making membrane perforation by exposure of cells to ultrasound, and it is an attractive method for realizing nonvirus gene transfection. A continuous or quasicontinuous wave is frequently used for this technique because a higher duty ratio gives higher efficiency of sonoporation. Addition of microbubbles during insonification greatly improves the efficiency of sonoporation, and, especially when microbubbles exist in the vicinity of the cells, ultrasound pulses from diagnostic ultrasound equipment can cause sonoporation. In this study, we examined sonoporation induced by single-shot pulsed ultrasound and the role of microbubbles in induction of cell membrane perforation. Bubble behavior and cell membrane damage were observed using a high-speed camera and light and scanning electron microscopes. Results of observation showed that mechanical stress induced by bubble motion could cause cell membrane perforation. We also studied repair of the perforation using a fluorescence microscope and found that the membrane of mammalian cells has the ability to reseal the perforation within several seconds. [Research partially supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan.]

8:40

4aBB3. Quantitative imaging of tumor blood flow with contrast ultrasound. Peter N. Burns, Raffi Karshafian, and John Hudson (Dept. Medical Biophys., 2075 Bayview Ave., Toronto ON, M4N 3M5, Canada)

The point at which a solid cancer develops its own blood supply marks the onset of malignant progression. This process, known as angiogenesis, makes oxygen and nutrients available for growth and provides a path for metastatic spread. Angiogenesis is not only of interest as a diagnostic hallmark of malignancy, but also as a target for new therapeutic strategies. Assessing antiangiogenic therapies noninvasively poses problems—flow velocities (<1 mm/s) and vessel diameters (<50 μm) are below resolutions of direct imaging. Vessels are disorganized without the tree-like structure of normal vasculature. We have investigated the potential role of

microbubble disruption-replenishment flow measurement in monitoring antivascular treatment of an animal tumor. The currently used monexponential model incorrectly considers the vasculature a perfect mixing chamber. Simple fractal models of the circulation provide a distribution of vessel diameters which, combined with the geometry of the disruption and detection beams, produce better models of replenishment following acoustic bubble disruption. These not only measure flow, but also predicts differences between organized and disorganized circulations, even with equal flow and vascular volume. If detectable, such differences might be used to characterize vascular organization below the resolution limit of an ultrasound image.

9:00

4aBB4. Dynamics of laser-trapped microbubbles. Hiroyuki Takahira (Dept. of Mech. Eng., Osaka Prefecture Univ., 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan)

A laser-trapping method is utilized for microbubbles. Bubbles of the order of 10 microns in diameter are trapped and manipulated successfully using a dry objective lens with large working distance. The growth or shrinkage of a laser-trapped microbubble and the merger of microbubbles are observed with a high-speed camera to investigate the influence of gas diffusion on the stability of microbubbles. Two kinds of equilibrium radii are found for shrinking microbubbles. The first one is related to the equilibrium surface concentration of surfactant. The other is related to the decrease of the surface tension due to the compression of the surface area at the maximum surfactant concentration. The simulations in which the dynamic surface tension is considered are in good agreement with the experiments. The laser trapping technique is also applied to the motion of a microbubble in a shear flow. It is shown that the bubble escapes from the laser trap being repelled by the optical force in the shear flow. There is overall agreement between the experiments and the simulations in which the buoyancy force, the fluid dynamic forces, and the optical force are taken into account.

9:20

4aBB5. Novel methods of micro-object trapping by acoustic radiation force. Yoshiki Yamakoshi (1-5-1 Tenjin-cho, Kiryushi, Gunma 376-8515 Japan, yamakosi@el.gunma-u.ac.jp)

It is expected that micro object trapping by acoustic radiation force is a useful method in future drug delivery systems in order to concentrate the payloads at desired position. In this paper, two novel methods of micro object trapping are presented. First method is micro object trapping by seed bubbles. This method uses seed bubbles, which have higher sensitivity to the ultrasonic wave, in order to trap micro objects, which are difficult to trap by conventional methods due to low volumetric oscillation under the ultrasonic wave. The Bjerkne's force, which is produced by a secondary wave radiated from the seed bubbles, traps the target objects making bi-layer seed bubbletarget object mass. The Second method is micro bubble trapping by bubble nonlinear oscillation. Two ultrasonic waves with different frequencies (pumping and control waves) are introduced simultaneously. The frequency of the control wave is set to a harmonic frequency of the pumping wave. If the bubbles flow into the cross area of two waves, nonlinear oscillation by high intensity pumping wave generates the Bjerkne's force, producing multiple traps with narrow separation along the control wave propagation direction. In order to demonstrate these methods, experiments using an ultrasonic wave contrast agent are shown.

9:40

4aBB6. Mechanical properties of HeLa cells at different stages of cell cycle by time-resolved acoustic microscope. Pavel V. Zinin (School of Ocean and Earth Sci. and Technol., Univ. of Hawaii, 2525 Correa Rd., Honolulu, HI 96822-2219), Eike C. Weiss, Pavlos Anastasiadis, and Robert M. Lemor (Fraunhofer Inst. for Biomed. Technol., St. Ingbert, Germany)

Scanning acoustic microscopy (SAM), particularly time-resolved acoustic microscopy, is one of the few techniques for study of the mechanical properties of only the cell's interior, cytosol and nucleus. Unfortunately, time-resolved acoustic microscopes typically do not provide sufficient resolution to study the elasticity of single cells. We demonstrate that the high-frequency, time-resolved acoustic microscope developed at the Fraunhofer Institute for Biomedical Technology (IBMT), Germany, is capable of imaging and characterizing elastic properties of micron size structures in cell's cytoskeleton with a theoretical resolution limit of 10 m/s for sound speed measurements. Measurements were performed on cells of the HeLa cell line derived from human cervics carcinoma. SAM measurements of the sound speed of adherent HeLa cells at different states of the cell cycle were conducted. They yielded an average value of 1540 m/s. B-Scan images of HeLa cells at different states of the cell cycle show distinct patterns inside the cell. A method for estimating sound attenuation inside HeLa cells is outlined as such a method is critical for the determination of a cell's viscoelasticity. [Work supported by Alexander von Humboldt Foundation and the European Framework Program 6, Project "CellProm."]

10:00−10:10   Break

10:10

4aBB7. Assessment of shock wave lithotripters via cavitation potential. Jonathan I. Iloreta, Andrew J. Szeri (UC Berkeley, 6119 Etcheverry Hall, M.S. 1740, Berkeley, CA 94720-1740), Yufeng Zhou, Georgii Sankin, and Pei Zhong (Duke Univ., Durham, NC 27708-0300)

An analysis of bubbles in elastic media has been made in order to characterize shock wave lithotripters by gauging the potential for cavitation associated with the lithotripter shock wave (LSW). The method uses the maximum radius achieved by a bubble subjected to a LSW as the key parameter that defines the potential damage a lithotripter could cause at any point in the domain. The maximum radius is determined by an energy analysis. A new index—similar in spirit to the Mechanical Index of Holland and Apfel for diagnostic ultrasound—is proposed for use in gauging the likelihood of cavitation damage.

10:30

4aBB8. Formation of water pore in a bilayer induced by shock wave: Molecular dynamics simulation. Kenichiro Koshiyama, Takeru Yano, Shigeo Fujikawa (Lab. of Adv. Fluid Mech., Hokkaido Univ., Sapporo 060-8628, Japan, koshi@ring-me.eng.hokudai.ac.jp), and Tetsuya Kodama (Tohoku Univ., Aobaku, Sendai 980-8575, Japan)

The irradiation of a shock wave or ultrasound with micro-bubbles has the potential to make transient pores on cell membranes. Although such pores are believed to contribute to the molecular delivery thorough the membrane, the detailed mechanisms of the pore formation with shock waves and the subsequent molecular delivery through the pores into cells are still unclear. To investigate the mechanism at a molecular level, the molecular dynamics simulations of the interaction of the shock wave with a lipid bilayer are conducted. The water penetration into the hydrophobic region by the shock wave is observed in picoseconds. As a next step, structural changes of the bilayer containing water molecules in the hydrophobic region are investigated. The water pore is formed in 3 ns when the large number of water molecules is inserted. The lifetime of the water pore is more than 70 ns. The radius of the water pore is ca. 1.0 nm, which is three times larger than the Stoke's radius of a typical anticancer drug (5FU). Finally, the diffusion of the anticancer drug in the water pore is investigated.

10:45

4aBB9. Ultrasonic spore lysis and the release of intracellular content in a microfluidic channel. Oana C. Marina, Michael D. Ward, John M. Dunbar, and Gregory Kaduchak (MPA-11, Los Alamos Natl. Lab., P.O. Box 1663, MS D-429, Los Alamos, NM, 87545)

Ultrasonic lysis of suspended spores in a microfluidic channel is a promising alternative to conventional spore disruption techniques that include bead beating as the spore lysis gold standard. Our overall research goal is to obtain an automated detection system with complete sample preparation and lysis steps in a microfluidic channel. Previously, much work in this area has focused on organism viability rather than the release of intracellular material. Our research focuses on quantifying the amount of intracellular content (e.g., DNA, proteins, etc.) that is released by acoustic lysis for detection by the sensor. Elucidating the efficacy of acoustics on the release of intracellular material requires reliable methods to quantify the released intracellular content (nucleic acids and proteins). The device used for lysing spores consists of a microfluidic chamber with one acoustically active wall. The chamber depths are in the range of 100–200 m. Channels tested in the 70-kHz to 1-MHz frequency range show that the efficiency of intracellular release depends on the operating frequency of the device and the properties (concentration, composition) of the spore suspensions. Experimental results on viability and released intracellular content are discussed. [Work supported by LANL LDRD.]

11:00

4aBB10. The correlation between cavitation noise power and bubble-induced heating in high-intensity focused ultrasound. Caleb H. Farny, Tianming Wu, R. Glynn Holt, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, cfarny@bu.edu)

It has been established that inertial cavitation is responsible for elevated heating during high-intensity focused ultrasound (HIFU) application for certain intensity regimes. The contribution of bubble-induced heating can be an important factor to consider, as it can be several times that expected from absorption of the primary ultrasound energy. Working in agar-graphite tissue phantoms with a 1.1-MHz HIFU transducer, an embedded type-E thermocouple, and a 15-MHz passive cavitation detector (PCD), the temperature and cavitation signal near the focus were measured for 5-s continuous wave HIFU insonations. The measured temperature was corrected for heating predicted from the primary ultrasound absorption and the transient thermocouple viscous heating artifact to isolate the temperature rise from the bubble activity. We have found that the temperature rise induced from the bubble activity correlates well with the instantaneous cavitation noise power as indicated by the mean square voltage output of the PCD. The results suggest that careful processing of the cavitation signals could serve as a proxy for measuring the heating contribution from inertial cavitation. [Work supported by the Dept. of the Army (Award No. DAMD17-02-2-0014) and the Center for Subsurface Sensing and Imaging Systems (NSF ERC Award No. EEC-9986821).]

11:15

4aBB11. Membrane permeabilization of adherent cells with laser-induced cavitation bubbles. Rory Dijkink, Claus-Dieter Ohl (Phys. of Fluids, Univ. of Twente, Postbus 217, 7500 AE Enschede, The Netherlands), Erwin Nijhuis, Sèverine Le Gac (Univ. of Twente, 7500 AE Enschede, The Netherlands), and Istvàn Vermes (Medical Spectrum Twente Hospital Group, 7500 KA Enschede, The Netherlands)

Strongly oscillating bubbles close to cells can cause the opening of the cell's membrane, thus to stimulate the uptake of molecules from the exterior. However, the volume oscillations of bubbles induce complex fluid flows, especially when bubble-bubble interaction takes place. Here, we report on an experiment where a single cavitation bubble is created close to a layer of adherent HeLa cells. The interaction distance between the bubble and the cell layer is controlled by adjusting the focus of the pulsed laser light, which creates the cavitation bubble. The dynamics of the bubble and the cells is recorded with high-speed photography. The poration of the cells is probed with different fluorescent stains to distinguish viable and permanent poration and programmed cell death (apoptosis). Quantitative data are presented as a function of the radial distance from the stagnation point. Our main finding is the importance of the asymmetrical collapse and the high-speed jet flow: After impact of the jet onto the substrate a strong boundary layer flow is responsible for shearing the cells.

11:30

4aBB12. Antitumor effectiveness of cisplatin with ultrasound and nanobubbles. Tetsuya Kodama, Yukiko Watanabe, Kiyoe Konno, Sachiko Horie (Res. Organization, Tohoku Univ., 2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi 980-8575, Japan), Atsuko Aoi (Tohoku Univ., Sendai 980-8575, Japan), Geroges Vassaux (Bart's and The London School of Medicine and Dentistry, UK), and Shiro Mori (Tohoku Univ. Hospital, Sendai 980-8575, Japan)

The potentiation of antitumor effect of cis-diamminedichloroplatinum (II), cisplatin, with ultrasound (1 MHz, 0.6 MPa) and lipid-shelled nanobubbles *in vitro* (EMT6, C26, MCF7, A549) and *in vivo* on s.c. tumor in mice (HT29-expressing luciferase) were evaluated. *In vitro* and *in vivo* antitumor effects were measured by an MTT assay and a real-time *in vivo* imaging, respectively. The effective antitumor effect was seen both *in vitro* and *in vivo* when ultrasound and nanobubbles were used, while other treatment groups with cisplatin with ultrasound did not show the effectiveness. The antitumor effect was not attributed to necrosis but apoptosis, which was confirmed by increase in the activity of the pro-apoptosis signal caspase-3 and Bax. In conclusion, the combination of ultrasound and nanobubbles with cisplatin is an effective chemotherapy of solid tumors and may prove useful in clinical application.

11:45

4aBB13. Sonoporation by single-shot pulsed ultrasound with microbubbles—Little effect of sonochemical reaction of inertial cavitation. Kengo Okada, Nobuki Kudo, and Katsuyuki Yamamoto (Grad. School of Information Sci. and Technol., Hokkaido Univ., Kita 14 Nishi 9, Kita-ku, Sapporo 060-0814, Japan)

Sonoporation is a technique for introducing large molecules into a cell by exposure to ultrasound, and it has a potential application for gene transfection. Although continuous-wave ultrasound is generally used for this technique, we have been using single-shot pulsed ultrasound with microbubbles. To determine the contribution of the sonochemical effect of

4a FRI. AM

inertial cavitation under the condition of single-shot exposure, we compared rates of cell membrane damage in the presence and absence of a free radical scavenger (cysteamine, 5 mM). Cells with microbubbles in their vicinity were exposed to pulsed ultrasound of 1.1 MPa in negative peak pressure under microscopic observation, and the numbers of total and damaged cells in the view field were counted. The damage rates were $8.1\pm4.0\%$ and $10.3\pm6.3\%$ in the presence ($n=17$) and absence (n =25) of the scavenger, respectively, and the average number of total cells was 772 ± 285. Since there was no significant difference, we concluded that the cell membrane damage observed in our exposure condition was not caused by the sonochemical effect but by the mechanical effect of inertial cavitation. [Research was supported by a grant-in-aid for scientific research from the Ministry of Education, Science, Sports and Culture, Japan.]

FRIDAY MORNING, 1 DECEMBER 2006

OAHU ROOM, 8:00 TO 11:50 A.M.

Session 4aEA

Engineering Acoustics and ASA Committee on Standards: Developments in Microphones: Calibrations, Standards, and Measures

George S. K. Wong, Cochair

National Research Council, Inst. for National Measurement Standards, 1500 Montreal Rd., Ottawa, Ontario K1A 0R6, Canada

Masakasu Iwaki, Cochair

NHK Science and Technology Research Labs., 1-10-11 Kinuta, Setagaya-ku, Tokyo 157-8510, Japan

Chair's Introduction—8:00

Invited Papers

8:05

4aEA1. Current developments at the National Institute for Standards and Technology in pressure calibration of laboratory standard microphones. Victor Nedzelnitsky, Randall P. Wagner, and Steven E. Fick (National Inst. of Standards and Technol. [NIST], 100 Bureau Dr., Stop 8220, Gaithersburg, MD 20899-8220, victor.nedzelnitsky@nist.gov)

Current research effort aims at improving the apparatus and methods for determining the pressure sensitivities of IEC types LS1Pn and LS2aP laboratory standard microphones. Among the improvements that are being systematically incorporated in an evolving test bed is the capability to operate at adjustable power line frequencies other than the usual 60 Hz. Suitable choices of line frequency relative to frequencies of calibration and adjustable bandpass filter characteristics can be used to improve the signal-to-noise ratios of measurements performed near the usual line frequency and its first few harmonics. This can enable the use of relatively large volume couplers for which uncertainties in microphone front cavity volume and equivalent volume, capillary tube effects, and heat conduction corrections have a lesser influence than they have for small-volume couplers. Another improvement aims to control and to stabilize the ambient static pressure during microphone calibrations, to reduce or eliminate the effects of barometric pressure fluctuations on these calibrations.

8:25

4aEA2. Free-field reciprocity calibration of laboratory standard (LS) microphones using a time selective technique. Knud Rasmussen and Salvador Barrera-Figueroa (Danish Primary Lab. of Acoust. (DPLA), Danish Fundamental Metrology, Danish Tech. Univ., Bldg. 307, 2800 Kgs., Lyngby, Denmark)

Although the basic principle of reciprocity calibration of microphones in a free field is simple, the practical problems are complicated due to the low signal-to-noise ratio and the influence of cross talk and reflections from the surroundings. The influence of uncorrelated noise can be reduced by conventional narrow-band filtering and time averaging, while correlated signals like cross talk and reflections can be eliminated by using time-selective postprocessing techniques. The technique used at DPLA overcomes both these problems using a B&K Pulse analyzer in the SSR mode (steady state response) and an FFT-based time-selective technique. The complex electrical transfer impedance is measured in linear frequency steps from a few kHz to about three times the resonance frequency of the microphones. The missing values at low frequencies are estimated from a detailed knowledge of the pressure sensitivities. Next an inverse FFT is applied and a time window around the main signal is used to eliminate cross talk and reflections. Finally, the signal is transformed back to the frequency domain and the free field sensitivities calculated. The standard procedure at DPLA involves measurements at four distances and the repeatability of the calibrations over time is within ±0.03 dB up to about 1.5 times the resonance frequency of the microphones.

3232   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3232**

4aEA3. Microphone calibration by comparison. George S. K. Wong (Acoust. Standards, Inst. for Natl. Measurement Standards, National Res. Council Canada, Ottawa, ON K1A 0R6, Canada)

The absolute method of microphone calibration by the reciprocity method (IEC 61094-2 1992) provides the highest accuracy of approximately 0.04 to 0.05 dB, and the procedure requires three changes of microphone in the "driver-receiver combination" that needs approximately 1 to 2 days. The system capital cost is relatively high. The NRC interchange microphone method for microphone calibration by comparison has been adopted internationally by the International Electrotechnical Commission as IEC 61094-5 (2001-10). With this method, the test microphone is compared with a reference microphone calibrated by the reciprocity method and the procedure requires approximately 3 h. The uncertainty of the comparison method is between 0.08 to 0.09 dB, which satisfies most industrial needs.

4aEA4. Development of a laser-pistonphone for an infrasonic measurement standard. Ryuzo Horiuchi, Takeshi Fujimori, and Sojun Sato (Natl. Metrology Inst. of Japan (NMIJ), AIST, Tsukuba Central 3, 1-1-1 Umezono, Tsukuba, 305-8563, Japan)

Acoustical standards for audio frequencies are based on pressure sensitivities of laboratory standard microphones calibrated using a coupler reciprocity technique. There is a growing need to extend the frequency range downward for reliable infrasonic measurement. The reciprocity technique, however, has limitations on low-frequency calibration (1–20 Hz) because signal-to-noise ratio deteriorates and a sound leak occurs from capillary tubes that equalize the static pressure inside and outside of the coupler. These factors rapidly increase the measurement uncertainty as the frequency is lowered. NMIJ has therefore recently developed a laser-pistonphone prototype, which enables precise calibration of microphones at low frequencies. Compared with the reciprocity technique, the laser-pistonphone produces a higher sound pressure within a cavity by the sinusoidal motion of a piston and has a significantly improved signal-to-noise ratio. Sound pressure is calculated from the piston displacement, which is determined via a Michelson interferometer. A test microphone is inserted into the cavity, exposed to the sound pressure, and its open-circuit voltage is measured. Static pressure equalization is realized through the gap between the piston and its guide. Careful design of the dimensions and relative position of the cavity and piston minimizes sound leakage and friction between them.

4aEA5. Anechoic measurements of particle-velocity probes compared to pressure gradient and pressure microphones. Wieslaw Woszczyk (CIRMMT, McGill Univ., 555 Sherbrooke St. West, Montreal, QC, Canada H3A 1E3, wieslaw@music.mcgill.ca), Masakazu Iwaki, Takehiro Sugimoto, Kazuho Ono (NHK Sci. & Tech. Res. Labs., Setagaya-ku, Tokyo 157-8510, Japan), and Hans-Elias de Bree (R&D Microflown Technologies)

Microflown probes are true figure-of-eight-pattern velocity microphones having extended response down to below the lowest audible frequencies, low noise, and high output. Unlike pressure-gradient microphones, velocity probes do not measure acoustic pressure at two points to derive a pressure gradient. When particle velocity is present, acoustical particle velocity sensors measure the temperature difference of the two closely spaced and heated platinum wire resistors, and quantify particle velocity from the temperature measurement. Microflown probes do not require a membrane and the associated mechanical vibration system. A number of anechoic measurements of velocity probes are compared to measurements of pressure-gradient and pressure microphones made under identical acoustical conditions at varying distances from a point source having a wide frequency response. Detailed measurements show specific response changes affected by the distance to the source, and focus on the importance of transducer calibration with respect to distance. Examples are given from field work using microflown probes to record acoustic response of rooms to test signals. The probe's cosine directional selectivity can be used to change the ratio between early reflections and the diffuse sound since only the $\frac{1}{3}$ of the power in the diffuse sound field is measured with the particle velocity probe.

4aEA6. Sensitivity change with practical use of electret condenser microphone. Yoshinobu Yasuno (Panasonic Semiconductor Device Solutions Co., Ltd. 600, Saedo-cho, Tsuzuki-ku, Yokohama, 224-8539, Japan) and Kenzo Miura (Panasonic Mobile Commun. Eng. Co., Ltd., Yokohama, Japan)

Dr. Sessler and Dr. West invented the electret condenser microphone (ECM) in 1966. It has since been applied in various ways as a sound input device. The ECM has become an important component as a microphone for communications because of its stable sensitivity frequency characteristic. Materials and production methods have been improved continually up to the present. In particular, the ECM reliability is based on the electret's stability. For that reason, the electret surface charge decay is the main factor in ECM sensitivity degradation. This study analyzed the changes of an ECM preserved for 28 years in the laboratory and actually used for an outdoor interphone unit for 29 years. The change of diaphragm stiffness and electret surface voltage were compared with the evaluation result of a heat-acceleration test and verified. A degradation estimate of sensitivity change of ECM was performed. Regarding the life of the electret predicted in the report of former study [K. Miura and Y. Yasuno, J. Acoust. Soc. Jpn. (E) **18**(1), 29–35 (1997)], the validity was verified using actual data from this long-term observation.

10:20

4aEA7. Development of a small size narrow directivity microphone. Masakazu Iwaki, Kazuho Ono, Takehiro Sugimoto   (NHK Sci. & Technol. Res. Labs., 1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8510, Japan), Takeshi Ishii, and Keishi Inamaga   (Sanken Microphone Co., Ltd, 2-8-8 Ogikubo, Suginami-ku, Tokyo, 167-0051, Japan)

We developed a new microphone that has very sharp directivity even in the low-frequency band. In an ordinary environment of sound pick-up, the energy of background noise is distributed mainly in frequencies lower than 1000 Hz. In such frequencies, a typical cardioid microphone has the directivity pattern close to that of omni-cardiod. Consequently, it is difficult to pick up the objective sounds clearly from background noises. To suppress the noises with very small level, the directivity pattern should be also sharpened in the low-frequency band. In this report, we describe a new method to sharpen directivity for the low-frequency band. The method requires three microphone capsules. One capsule is the main microphone with a very short acoustic pipe. The others compose a second-order gradient microphone to cancel the signal that comes from behind the main microphone. A special feature of this microphone is to control a dip frequency of behind sensitivity without changing the frequency response of the front sensitivity.

10:40

4aEA8. Two-wafer bulk-micromachined silicon microphones. Jianmin Miao and Chee Wee Tan   (Micromachines Ctr., School of Mech. and Aerosp. Eng., Nanyang Technolog. Univ., 50 Nanyang Ave., Singapore 639798, mjmmiao@ntu.edu.sg)

A two-wafer concept is proposed for silicon microphone manufacturing by using bulk-micromachining and wafer bonding technologies. Acoustical holes of the backplate in one wafer are micromachined by deep reactive ion etching and the diaphragm on another wafer is created by wet-chemical etching. The two wafers are then bonded together to form silicon condenser microphones. In order to minimize the mechanical-thermal noise and increase the sensitivity within the required bandwidth, an analytical model based on Zuckerwar's equations has been developed to find the optimum location of the acoustical holes in the backplate of microphones. According to our study, this analytical modeling has shown excellent agreement between the simulated and measured results for the B&K MEMS microphone. Silicon condenser microphones have been further optimized in terms of the air gap, number and location of acoustical holes to achieve the best performance with a low polarization voltage, and easy fabrication for possible commercialization. Details of analytical modeling, fabrication, and measurement results will be presented.

11:00

4aEA9. Infrasound calibration of measurement microphones. Erling Frederiksen   (Bruel & Kjaer, Skodsborgvej 307, 2850 Naerum, Denmark, erlingfred@bksv.com)

Increasing interest for traceable infrasound measurements has caused the Consultative Committee for Acoustics, Ultrasound and Vibration (CCAUV) of BIPM to initiate a key comparison calibration project (CCAUV.A-K2) on pressure reciprocity calibration down to 2 Hz. Ten national metrology institutes, including the Danish Primary Laboratory of Acoustics (DPLA), take part in this project. In addition DPLA has started its own infrasound calibration project, which is described in this paper. The purposes of this project are verification of the CCAUV results and development of methods for calibration of general types of measurement microphone between 0.1 and 250 Hz. The project includes the design of an active comparison coupler, an experimental low-frequency reference microphone, and new methods for its frequency response calibration. One method applies an electrostatic actuator and requires a low-pressure measurement tank, while the other requires an additional microphone, whose design is closely related to that of the reference microphone that is to be calibrated. The overall calibration uncertainty ($k=2$) for ordinary measurement microphones is estimated to less than 0.05 dB down to 1 Hz and less than 0.1 dB down to 0.1 Hz, if the reference is calibrated in the latter mentioned way, i.e., by the related microphones method.

Contributed Papers

11:20

4aEA10. Free-field calibration of 1/4 inch microphones for ultrasound by reciprocity technique. Hironobu Takahashi, Takeshi Fujimori, Ryuzo Horiuchi, and Sojun Sato   (Natl. Metrology Inst. of Japan, AIST, Tsukuba Central 3, 1-1-1 Umezono, Tsukuba, 305-8563 Japan)

Recently, equipment that radiates ultrasound radiation at frequencies far beyond the audible range is increasing in our environment. Such electronic equipment has switching regulators or inverter circuits, and many devices are unintended sources of ultrasound radiation. However, the effects of airborne ultrasound on human hearing and the human body have not been well investigated. To estimate the potential damage of airborne ultrasound radiation quantitatively, it is necessary to establish an acoustic standard for airborne ultrasound because the standard is a basis of acoustic measurement. With the intention of establishing a standard on airborne ultrasound, a free-field calibration system with an anechoic chamber was produced. The principle of free-field calibration techniques is introduced in this presentation. Type WS3 microphones (B&K 4939) were calibrated in the system to examine the calibration ability to be achieved. Results

showed that it can calibrate a microphone from 10 to 100 kHz with dispersion of less than 1 dB. In addition, the effects that were dependent on the uncertainty of the calibration are discussed based on those results.

11:35

4aEA11. An environmentally robust silicon diaphragm microphone. Norihiro Arimura, Juro Ohga   (Shibaura Inst. of Technol.,3-7-5 Toyosu, Koto-ku, Tokyo, 135-8548, Japan), Norio Kimura, and Yoshinobu Yasuno (Panasonic Semiconductor Device Solutions Co., Ltd., Saedo-cho, Tsuzuki-ku, Yokohama, Japan)

Recently, many small microphones installed in cellular phones are the electret condenser microphones (ECMs) that contain an organic film diaphragm. Although FEP of fluorocarbon polymer is generally used as the electret material, silicon dioxide is also used. Recently ECMs have been made small and thin while maintaining the basic sound performance according to the market demand. In addition, environment tests and the reflow soldering mounting process have been adjusted to meet market requirements. On the other hand, the examination satisfied the demand as

the high temperature resistance was insufficient. This paper depicts an examination and a comparison of conventional ECM with the experimental model, a silicon diaphragm condenser microphone produced using the MEMS method. The silicon diaphragm satisfies high-temperature resis-tance and stable temperature characteristics because of its very small co-efficient of linear expansion and it is measured total harmonic distortion (THD) on high pressure sound. Finally, it will be able to be used in high temperature and high pressure sound conditions in the future.

FRIDAY MORNING, 1 DECEMBER 2006   IAO NEEDLE/AKAKA FALLS ROOM, 8:15 TO 11:15 A.M.

Session 4aMU

Musical Acoustics: Music Information and Communication

Bozena Kostek, Cochair
Gdansk Univ. of Technology, Multimedia Systems Dept., Narutowicza 11- 12, 80-952 Gdansk, Poland

Masuzo Yanagida, Cochair
Doshisha Univ., Dept. of Information Science and Intelligent Systems, 1-3 Tatara-Miyakodani, Kyo-Tanabe, Kyoto 610-0321, Japan

Invited Papers

8:15

4aMU1. Introduction of the Real World Computing music database. Masataka Goto (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, m.goto@aist.go.jp)

This paper introduces the *RWC (Real World Computing) Music Database*, a copyright-cleared music database that is available to researchers as a common foundation for research. Shared databases are common in other research fields and have contributed importantly to progress in those fields. The field of music information processing, however, has lacked a common database of musical pieces and a large-scale database of musical instrument sounds. The RWC Music Database was therefore built in fiscal 2000 and 2001 as the world's first large-scale music database compiled specifically for research purposes. It contains six original collections: the *Popular Music Database* (100 pieces), *Royalty-Free Music Database* (15 pieces), *Classical Music Database* (50 pieces), *Jazz Music Database* (50 pieces), *Music Genre Database* (100 pieces), and *Musical Instrument Sound Database* (50 instruments). To address copyright issues, all 315 musical pieces were originally composed, arranged, or performed, and all instrumental sounds were originally recorded. The database has already been distributed to more than 200 research groups and is widely used. In addition, a continuous effort has been undertaken to manually annotate a set of music-scene descriptions for the musical pieces, called *AIST Annotation*, which consists of the beat structure, melody line, and chorus sections.

8:35

4aMU2. Japanese traditional singing on the same lyrics. Ichiro Nakayama (Osaka Univ. of Arts, 469, Higashiyama, Kanan-cho, Minami-Kawachi-gun, Osaka, 585-8555 Japan) and Masuzo Yanagida (Doshisha Univ., Kyo-Tanabe, 610-0321 Japan)

Described is a database of Japanese traditional singing together with supplementary recording of Bel Canto for comparative studies. Singing sounds and spoken speech by the same singers are recorded in pair to form the body of the database. This database covers most of genres of Japanese traditional singing, such as Shinto prayers, Buddist prayers, Nor, Kyogen, Heikyoku, Sokyoku, Gidayu-bushi, Kabuki, Nagauta, Tokiwazu, Kiyomoto, Itchu-bushi, Shinnai, Kouta, Zokkyoku, Rokyoku, Shigin, Ryukyu-clasico, Goze-uta, etc. All the sounds were recorded in anechoic chambers belonging to local institutions, mainly in Osaka and Tokyo, asking 78 professional singers including 18 "Living National Treasures" to sing as informants. The most important point of this database is that an original lyric especially prepared for this recording is commonly used to make comparative studies easy. All the subjects are asked to sing the common lyrics in their own singing styles. Shown here are comparisons of formant shifts in vowels from ordinary speaking to singing for some singers, and comparison of temporal features of fundamental frequency between Japanese traditional singing and Western Bel Canto. [Work supported by the Academic Frontier Project, Doshisha University.]

8:55

4aMU3. Computational intelligence approach to archival musical recordings. Andrzej Czyzewski, Lukasz Litwic, and Przemyslaw Maziewski (Gdansk Univ. of Technol., Narutowicza 11/12, 80-952 Gdansk, Poland)

An algorithmic approach to wow defect estimation in archival musical recordings is presented. The wow estimation is based on the simultaneous analysis of many sinusoidal components, which are assumed to depict the defect. The rough determination of sinusoidal components in analyzed musical recording is performed by standard sinusoidal modeling procedures employing magnitude and phase spectra analysis. Since archival recordings tend to contain distorted tonal structure, the basic sinusoidal modeling approach is often found insufficient, resulting in audible distortions in the restored signal. It is found that the standard sinusoidal modeling approach is prone to errors, especially when strong frequency or amplitude variations of sinusoidal components occur. It may result in gaps or inappropriately matched components, leading to incorrect estimation of the wow distortion. Hence, some refinements to sinusoidal component analysis including interpolation and extrapolation of tonal components are proposed. As it was demonstrated in experi-

4a FRI. AM

ments, due to the nonlinear nature of wow distortion, the enhancement of sinusoidal analysis can be performed by means of a neural network. The paper demonstrates implemented algorithms for parasite frequency modulation in archival recordings together with obtained results. [Work supported by the Commission of the European Communities, within the Integrated Project No. FP6-507336: PRESTOSPACE.]

9:15

4aMU4. Music information retrieval seen from the communication technology perspective. Bozena Kostek (Gdansk Univ. of Technol., Narutowicza 11/12, PL-80-952 Gdansk, Poland)

Music information retrieval (MIR) is a multidiscipline area. Within this domain one can see various approaches to musical instrument recognition, musical phrase classification, melody classification (e.g., query-by-humming systems), rhythm retrieval, high-level-based music retrieval such as looking for emotions in music or differences in expressiveness, and music search based on listeners' preferences. One may also find research that tries to correlate low-level descriptor analysis to high-level human perception. Researchers from musical acoustics, musicology, and music domains on one side, and communication technology on the other side, work together within this area. This may foster a framework for broader and deeper comprehension of contributions from all these disciplines and, in addition, translate the automated access to music information, gathered in various forms around the World Wide Web, as a fully understandable process to all participants regardless of their background. The semantic description is becoming a basis of the next Web generation. Several important concepts have been introduced recently by the researchers associated with the MIR community with regard to semantic data processing including techniques for computing with words. In this presentation some aspects related to MIR are briefly reviewed in the context of possible and actual applications of ontology-based approach to this domain.

9:35

4aMU5. Accompaniment included song waveform retrieval based on framewise phoneme recognition. Yuichi Yaguchi and Ryuichi Oka (Univ. of Aizu, Tsuruga, Ikkimachi, Aizuwakamatsu, Fukushima, 965-8580 Japan)

A novel approach is presented for a retrieval method that is useful for waveforms of songs with accompaniment. Audio signals of songs have some different acoustical characteristics from speech signals. Furthermore, the length per mora of signals is longer than that of speech. Therefore, the authors suggest a sound retrieval system for application to musical compositions, including songs, that extracts framewise acoustical characteristics and uses a retrieval method for absorbing phoneme length. First, the system prepares two sets of phoneme identification functions that have corresponding order, but for which phoneme sets belong to different environments of accompaniment-included or accompaniment-reduced. Next, musical compositions are put into database and the query song wave converts a waveform to a label sequence using framewise phoneme recognition derived by Bayesian estimation that applies each phoneme identification function according to whether it is accompaniment-included or not. Finally, the system extracts an interval area, such as query data, from a database using spotting recognition that is derived using continuous dynamic programming (CDP). Retrieval method results agree well with earlier results [Y. Yaguchi and R. Oka, AIRS2005, LNCS3689, 503–509 (2005)] that applied the same musical composition set without accompaniment.

9:55–10:10   Break

10:10

4aMU6. Design of an impression-based music retrieval system. Kimiko Ohta, Tadahiko Kumamoto, and Hitoshi Isahara (NICT, Keihanna Kyoto 619-0289, Japan, kimiko@nict.go.jp)

Impression-based music retrieval helps users to find musical pieces that suit their preferences, feelings, or mental states from among a huge volume of a music database. Users are asked to select one or more pairs of impression words from among multiple pairs that are presented by the system and to estimate each selected pair on a seven-step scale to input their impressions into the system. For instance, if they want to locate musical pieces that will create a happy impression, they should check the radio button "Happy" in the impression scale: Very happy–Happy–A little happy–Neutral–A little sad–Sad–Very sad. A pair of impression words with a seven-step scale is called an impression scale in this paper. The system calculates the distance between the impressions of each musical piece in a user-specified music database and the impressions that are input by the user. Subsequently, it selects candidate musical pieces to be presented as retrieval results. The impressions of musical pieces are expressed numerically by vectors that are generated from a musical piece's pitch, strength, and length of every tone using n-gram statistics.

Contributed Papers

10:30

4aMU7. Automatic discrimination between singing and speaking voices for a flexible music retrieval system. Yasunori Ohishi, Masataka Goto, Katunobu Itou, and Kazuya Takeda (Grad. School of Information Sci., Nagoya Univ., Furo-cho 1, Chikusa-ku, Nagoya, Aichi, 464-8603, Japan, ohishi@sp.m.is.nagoya-u.ac.jp)

This paper describes a music retrieval system that enables a user to retrieve a song by two different methods: by singing its melody or by saying its title. To allow the user to use those methods seamlessly without changing a voice input mode, a method of automatically discriminating between singing and speaking voices is indispensable. We therefore first investigated measures that characterize differences between singing and speaking voices. From subjective experiments, we found that human listeners discriminated between these two voices with 70% accuracy for 200-ms signals. These results showed that even short-term characteristics such as the spectral envelope represented as MFCC can be used as a discrimination cue, while the temporal structure is the most important cue when longer signals are given. According to these results, we then developed the automatic method of discriminating between singing and speaking voices by combining two measures: MFCC and an $F0$ (voice pitch) contour. Experimental results with our method showed that 68.1% accuracy was obtained for 200-ms signals and 87.3% accuracy was obtained for 2-s signals. Based on this method, we finally built a music retrieval system that can accept both singing voices for the melody and speaking voices for the title.

10:45

4aMU8. Various acoustical aspects of an Asian (South) Indian classical music concert. M. G. Prasad (Dept. of Mech. Eng., Stevens Inst. of Technol., Hoboken, NJ 07030), V. K. Raman (Flautist, Germantown, MD 20874), and Rama Jagadishan (Edison, NJ 08820)

An Asian (South) Indian classical music concert is an integrated acoustical experience for both the audience and the player(s). A typical concert team, either vocal or instrumental, consists of a main vocalist (or an instrumentalist) accompanied by a violinist, up to three percussion instrument players, and a reference drone. The concert is comprised of many songs. Each song has two main parts, namely Alapana and Kriti. The Alapana is an elaboration of a raga (tune) and the Kriti refers to the lyrics of the song. The violinist actively follows and supports the main musician during the concert. The percussion player(s) are provided an opportunity to present a solo of their rhythmic skills. The players and the audience communicate emotionally and intellectually with each other. Elements such as aesthetics, rhythm, skill, and emotional aspects of the players are evaluated and appreciated by the audience. This talk will present various aspects of a concert that brings about an integrated and holistic experience for both the audience and the player(s). Some samples from live vocal and instrumental music concerts will be presented.

11:00

4aMU9. Musical scales, signals, quantum mechanics. Alpar Sevgen (Dept. of Phys., Bogazici Univ., Bebek 34342, Istanbul, Turkey)

Scales, being finite length signals, allow themselves to be treated algebraically: key signatures are related to the "ring" property of the scale labels; cyclically permuted scales and their mirror images have the same number of sharps and flats; and complementary scales (like major and pentatonic) have their sharp and flat numbers exchanged. A search for minimum principles to select among all possible scales those employed in music yields two possibilities: (a) minimize total number of accidentals and (b) minimize frequency fluctuations in a scale. Either of these minimum principles helps filter those scales employed in music from the universe of all scales, setting up very different criteria than the harmonic ratios used by musicians. The notes of the scales employed in music seem to prefer to stay as far apart from each other as possible. Operators that step through the multiplet members of scales with N semitones form a complete set of operators together with those that step through their eigenvectors. The mathematics reveals the discrete Fourier transformations (DFT) and is identical to finite state quantum mechanics of N-level Stern-Gerlach filters worked out by J. Schwinger.

FRIDAY MORNING, 1 DECEMBER 2006

MAUI ROOM, 7:30 A.M. TO 12:15 P.M.

Session 4aNS

Noise and Architectural Acoustics: Soundscapes and Cultural Perception I

Brigitte Schulte-Fortkamp, Cochair

Technical Univ. Berlin, Inst. of Technical Acoustics, Secr TA 7, Einsteinufer 25, 10587 Berlin, Germany

Bennett M. Brooks, Cochair

Brooks Acoustics Corp., 27 Hartford Turnpike, Vernon, CT 06066

Invited Papers

7:30

4aNS1. Soundscape in the old town of Naples: Signs of cultural identity. Giovanni Brambilla (CNR Istituto di Acustica "O.M. Corbino" Via del Fosso del Cavaliere 100, 00133 Roma, Italy), Luigi Maffei, Leda De Gregorio, and Massimiliano Masullo (Second Univ. of Naples, 81031 Aversa (Ce), Italy)

Like all cities in Magna Grecia, the ancient Neapolis was built along three main parallel, tight, and straight streets called *decumani*. Since then and during the following centuries, commercial and handicraft activities, as well as social life, have been developed along these streets. The narrow ground rooms forced shopkeepers to occupy the main street to show their merchandise using vocal appeals to magnify their product, and handicrafts to work directly on the street (hammering, sawing, etc.). Music artists had their performance on the streets too. The soundscape in the area was a strong symbol of the Neapolitan cultural identity. Nowadays *decumani* have kept the main features of the past but some of these are overrun by road traffic. To investigate in which way the traffic noise has modified the soundscape perception and cultural identity, sound walks were registered during day and night time. A number of residents were interviewed and laboratory listening tests were carried out. Despite the congested urban environment and high sound levels, preliminary results have shown that part of the residential population is still able to identify the soundscape more related to Neapolitan historical identity.

7:50

4aNS2. Soundscape design in public spaces: Concept, method, and practice. Hisao Nakamura-Funaba and Shin-ichiro Iwamiya (Kyushu Univ., 4-9-1.Shiobaru, Minami-ku, Fukuoka 815-8540, Japan)

Soundscape design of public spaces necessitates consideration of whether a space has important meaning for the user. It is essential to imagine an ideal sound environment of that space. We designed an actual soundscape from the viewpoint of the environment, information, and decoration. In many cases, producing some silence in an environment becomes the first step of soundscape design. There is neither a special technology nor a technique when designing. It merely requires use of a general technology and techniques concerning sound according to location. A key point is knowledge of know how to coordinate these technologies and techniques. For instance, silence was made first at the renewal project of Tokyo Tower observatory through cancellation of its commercial and call broadcasting functions and installation of sound-absorbing panels to the ceiling. Next, suitable and small sounds were added at various points. Guests can take time to enjoy viewing as a result.

4aNS3. The daily rhythm of soundscape. Brigitte Schulte-Fortkamp (TU-Berlin, Einsteinufer 25 TA 7, 10587 Germany, brigitte.schulte-fortkamp@tu-berlin.de) and Andr Fiebig (HEAD acoustics GmbH, 52134 Herzogenrath, Germany)

With respect to people's minds, soundscapes can be considered as dynamic systems characterized by the time-dependent occurrence of particular sound events embedded in specific environments. Therefore, an adequate evaluation of environmental noise will reflect the continually varying acoustical scenery and its specific perception. An acoustical diary shall provide information about the daily routine and subjectively perceived sound exposure of residents. It relies on cognitive and emotional aspects of perceiving and evaluating sounds. It gives insight into evaluating processes and their contextual parameters because of its spontaneous character. It includes and refers to long-term acoustic measurements. Simultaneous measurements for elaborate acoustical analyses will be taken outside the homes regarding the moments of essential sound events. The aim is to collect information about the daily rhythm regarding acoustic events, whereby the focus also should be placed on sound events that are well accepted to deepen the explication of data with respect to the analysis. Procedure and results will be discussed.

4aNS4. Ecological explorations of soundscapes: From verbal analysis to experimental settings. Daniele Dubois (CNRS, 11 rue de Lourmel, 75015 Paris, France) and Catherine Guastavino (McGill Univ., Montreal, QC H3A 1Y1, Canada)

Scientific studies rely on rigorous methods that must be adapted to the object of study. Besides integrating acoustic features, soundscapes as complex cognitive representations also have the properties of being global, meaningful, multimodal, and categorical. Investigating these specificities, new paradigms were developed involving linguistics and ecological psychology to complement the psychophysical approach: cognitive linguistic analyses of discourse to address semantic properties of soundscapes, and categorization tasks and distances from prototypes to investigate their cognitive organization. As a methodological consequence, experimental settings must be designed to ensure the ecological validity of the stimuli processing, (the ''realism'' evaluated from a psychological point of view, stimuli being processed as in a real-life situation). This point will be illustrated with perceptual evaluations of spatial auditory displays for soundscape reproduction. Data processing techniques should also take into consideration the intrinsic properties of the representations they account for. Examples of free-sorting tasks will be presented with measurements in terms of family resemblance of sets of properties defining categories rather than dimensional scales. New ways of coupling physical measurement and psychological evaluations will be presented in order to simulate or reproduce soundscapes in both a realistic and controlled manner for experimental purposes.

4aNS5. Artificial neural network models of sound signals in urban open spaces. Lei Yu and Jian Kang (School of Architecture, Sheffield Univ., Western Bank, Sheffield S10 2TN, UK)

Sound signals, known as foreground sounds, are important components of soundscape in urban open spaces. Previous studies in this area have shown that sound preferences are different according to social and demographic factors of individual users [W. Yang and J. Kang, J. Urban Des., 10, 69–88 (2005)]. This study develops artificial neural network (ANN) models of sound signals for architects at design stage, simulating subjective evaluation of sound signals. A database for ANN modeling has been established based on large-scale social surveys in European and Chinese cities. The ANN models have consequently been built, where individual's social and demographic factors, activities, and acoustic features of the space and sounds are used as input variables while the sound preference is defined as the output. Through the process of training and testing the ANN models, considerable convergences have been achieved, which means that the models can be applied as practical tools for architects to design sound signals in urban open spaces, taking the characteristics of potential users into account. Currently ANN models combining foreground and background sounds are being developed.

4aNS6. Describing soundscape and its effects on people where soundscape is understood as an expansion of the concept of noise engineering. Keiji Kawai (Grad. School of Sci. and Technol., Kumamoto Univ., 2-39-1 Kurokami, Kumamoto 860-8555, Japan, kawai@arch.kumamoto-u.ac.jp)

This study discusses how to describe sound environment and people in terms of ''soundscape'' as an expansion of the ''noise engineering.'' In the framework of the conventional study field of noise evaluation, typically, sound environments are represented by loudness-based indices such as A-weighted sound pressure levels, and the impact of sound environments on people is represented by annoyance response or some physiological metrics. In the case of soundscape studies, however, the description should be expanded beyond what has been used in noise engineering. This matter has already been frequently discussed, but it doesn't seem that much consensus has been achieved concerning it yet. With respect to the effects of sound environment on people, since the concept of soundscape focuses on personal and social meanings of environmental sounds including the historical or aesthetic contexts, the effects are considered to be represented not by a singular concept such as comfortableness or quietness, but by multiple dimensions of emotional and aesthetic concepts. Also, descriptions of sound environment should include some qualitative aspects, such as what types of sounds can be heard at what extent. In this paper, the methodology to describe human-soundscape relationships is discussed through a review of related studies.

9:30

4aNS7. The sound environmental education aided by automated bioacoustic identification in view of soundscape recognition.
Teruyo Oba (Natural History Museum & Inst., Chiba, 955-2 Aoba-cho, Chuo-ku, Chiba-shi, Chiba-ken 260-8682 Japan, oba@chiba-muse.or.jp)

From the 2003–2004 JST projects, where the automated bioacoustic identification device Kikimimi-Zukin was introduced to the nature observation and environmental studies, it was learned that the activities encouraged children to take notice of sounds, become aware of the sound environment, and gain an insight into the soundscape. Sounds are often riddles to us, and hearing is the process to find out the facts and causes. It is more important to let children obtain appropriate clues to differentiate sounds by hearing and thinking for themselves than give them an immediate answer. Program on the Strength of Hearing was formulated to challenge children and have them enjoy hearing to identify, sharing what they hear with others, and observing environment through sounds. Kikimimi-Zukin reinforced the program by a step-by-step guide through the hearing process of scanning, focusing, characterizing, associating with relevant factors, and judging the identity. The experience not only brought them confidence in hearing but incentive to study nature and environment. With Kikimimi-Zukin children collected recordings and relevant information. Using the sound database, the local singing map and three-dimensional sound map were prepared. They facilitated communication on the local sound environment among children and with adults, leading to realization of their inner soundscape.

9:50

4aNS8. Acoustic environmental problems at temporary shelters for victims of the Mid-Niigata Earthquake.
Koji Nagahata, Norio Suzuki, Megumi Sakamoto, Fuminori Tanba (Fukushima Univ., Kanayagawa 1, Fukushima City, Fukushima, 960-1296, Japan, nagahata@sss.fukushima-u.ac.jp), Shin-ya Kaneko, and Tetsuhito Fukushima (Fukushima Medical Univ., Fukushima, 960-1295, Japan)

An earthquake on 23 October 2004 inflicted heavy damage on the Mid-Niigata district. The earthquake isolated Yamakoshi village; consequently, all the village residents were forced to evacuate to temporary shelters in neighboring Nagaoka city for 2 months. Two types of temporary shelters were used: gymnasiums, and buildings with large separated rooms similar to community centers. A questionnaire survey and interviews ($N=95$) were conducted to elucidate problems of the living environment at the temporary shelters. This study analyzed acoustic environmental problems there. Noise-related problems were noted by 40 respondents (46.5%): they were the fifth most frequently cited environmental problems. Several serious complaints, e.g., general annoyance at the shelters and footsteps of refugees at night, were only indicated by respondents who had evacuated to the gymnasiums. However, some problems, e.g., the clamor of children, including crying babies and voices of other refugees, were indicated by respondents irrespective of the type of the shelters to which they had been evacuated. Therefore, buildings like community centers were more desirable for temporary shelters, at least from the perspective of noise problems.

10:10–10:30   Break

10:30

4aNS9. The burden of cardiovascular diseases due to road traffic noise.
Wolfgang Babisch (Dept. of Environ. Hygiene, Federal Environ. Agency, Corrensplatz 1, 14195 Berlin, Germany, wolfgang.babisch@uba.de) and Rokho Kim (WHO/EURO Ctr. for Environment and Health, 53113 Bonn, Germany)

Epidemiological studies suggest a higher risk of cardiovascular diseases, including high blood pressure and myocardial infarction, in subjects chronically exposed to high levels of road or air traffic noise. A new meta-analysis was carried out to assess a dose-response curve, which can be used for a quantitative risk assessment and to estimate the burden of cardiovascular disease attributable to environmental noise in European regions. Noise exposure was grouped according to 5 dB(A)-categories for the daytime outdoor average A-weighted sound pressure level, (L_{day}, 16 h: 6–22 h), which was considered in most studies. Information on night-time exposure (L_{night}, 8 h: 22–6 h or 23–7 h) was seldom available. However, approximations can be made with respect to L_{den} according to the European directive on the assessment and management of environmental noise. The strongest evidence of an association between community noise and cardiovascular endpoints was found for ischaemic heart diseases, including myocardial infarction and road traffic noise. The disability-adjusted life years lost for ischemic heart disease attributable to transport noise were estimated conservatively, assuming the same exposure patterns across the countries with an impact fraction 3% in the western European countries.

10:50

4aNS10. Soundscape, moderator effects, and economic implications.
Cay Hehner (Henry George School of Social Sci., 121 E. 30th St., New York, NY 10016, chehner.hengeoschool@att.net) and Brigitte Schulte-Fortkamp (TU-Berlin, Berlin, Germany)

Soundscape is considered with respect to moderator effects and the contribution of economics. It will be questioned whether soundscapes can work as a moderator concerning noise annoyance. As shown by the different investigations concerning soundscapes, a definition of the meaning of soundscapes is necessary. Evidently, the moderating effect of a given environment and its soundscape has to be discussed on three levels: (1) extension of factors that describe annoyance,(2) peculiar feature of burdensome noise contexts, and (3) discrepancies of the social and economic status of people living in areas where the rebuilding will change the quality of the area. It has to be determined and analyzed to what extent the Georgist method of resources taxation, as recently exemplified, e.g., in Alaska and in Wyoming, can be instrumental in funding soundscapes to moderate noise annoyance as it has been the case in funding free education and allowing the distribution of a citizen's dividend.

4a FRI. AM

4aNS11. Socio-cultural soundscape concepts to support policies for managing the acoustic environment. Michiko So Finegold, Lawrence S. Finegold (Finegold & So, Consultants, 1167 Bournemouth Court Centerville, OH 45459-2647, m-so@pb3.so-net.ne.jp), and Kozo Hiramatsu (Kyoto Univ., Sakyou-ku, Kyoto 606-8501 Japan)

In the past half-century, considerable effort has been invested in the academic, technological, and political arenas to achieve an adequate acoustic environment. Various national and international policy guidance documents have made reasonable progress in establishing a framework for a common approach to minimizing environmental noise, such as documents from various national Environmental Protection Agencies, the World Health Organization, and the European Union. Although these documents have provided useful information for global application, they only address minimizing the negative side of the acoustic environment (i.e., noise), they focus primarily on acoustics issues at the national or international level, and they still have not adequately considered implementation issues related to socio-cultural differences. To deal with the practical problems that exist in the acoustic environment in the context of different cultures, continuing research and new policy guidance are needed to address different local situations and in a variety of cultural contexts. The Soundscape approach has been developing tools for describing the acoustic environment at the local level to address both the positive and negative aspects of the acoustic environment. In this paper, the evolving interdisciplinary aspect of the Socio-Cultural Soundscape will be discussed and key topics for future work will be recommended.

Contributed Papers

11:30

4aNS12. Initial steps for the determination of environmental noise quality—The perception-related evaluation of traffic noise. Klaus Genuit, Sandro Guidati, Sebastian Rossberg, and Andr Fiebig (HEAD Acoust. GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

Directives call for actions against noise pollution and noise annoyance. But, how do we eliminate harmful effects including annoyance due to the exposure of environmental noise without understanding the perception and evaluation of environmental noise? How do we preserve environmental noise quality where it is good (Directive 2002/49/EC) without identifying descriptors for noise quality? Various soundscape approaches based on different methodologies have been developed in the past. But, the measurement procedures must be realizable without much effort in order to achieve acceptance from legislation. Therefore, procedures have to be developed that capture the complexity of human hearing, on the one hand, and are feasible with respect to economy and time, on the other hand. The European project Quiet City (6FP PL516420) is dealing with, among other aspects, vehicle pass-by noise as a typical environmental noise source and its evaluation. Results of the analysis based on subjective assessments and psychoacoustic analyses carried out with respect to the development of an annoyance index will be presented and discussed. Such an index will provide valuable information for effective improvement of noise quality. The final aim is to develop a descriptor valid for complete traffic noise scenarios predicting environmental noise quality adequately.

11:45

4aNS13. When objective permissible noise limits of a municipal planning process and a subjective noise ordinance conflict. Marlund Hale (Adv. Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065, mehale@aol.com)

In most communities, proposed new building projects are required to conform with planning, community development, zoning, and/or building and safety specifications and standards. In cases of allowable noise exposure and noise limits, where certain of these requirements are quite specific while others are purposefully vague, conflicts between residential and commercial neighbors can lead to extreme disagreement and needless litigation. This paper describes a recent situation occurring in an upscale beach community, the resulting conflict over existing noise sources that comply with the limits of the city planning and permitting process, and the interesting findings of the court following the civil and criminal litigation that followed. Some suggestions are given to avoid these conflicting policy situations.

Contributed Poster Paper

Poster paper 4aNS14 will be on display from 7:30 a.m. to 12:15 p.m. The author will be at the poster from 12:00 noon to 12:15 p.m.

4aNS14. The complexity of environmental sound as a function of seasonal variation. Hideo Shibayama (3-7-5 Koutou-ku Tokyo, 135-8548, Japan, sibayama@sic.shibaura-it.ac.jp)

Residential land is performed for a surrounding area of a suburban local city. As a result of urbanization, an area of rich natural environments became narrow. For the animals and plants for whom a river and a forest are necessary, it becomes difficult to live. Environmental sound produced by the tiny insects in this area is changing from year to year. Catching the conditions for the developmental observations and environmental preservation in natural environments, we continue to measure the environmental sound as the time-series data. We estimate the complexity for these waveforms of the measured environmental sound in the season when insects chirp and do not chirp. For estimation of the complexity, we evaluate by the fractal dimension of the environmental sound. Environmental sound in early autumn is mainly generated by insects in the grass and on the trees. And, the fractal dimension for the sound waveforms of chirping of insects is up to 1.8.

Session 4aPA

Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration: Sound Propagation in Inhomogeneous Media I

James G. Miller, Cochair

Washington Univ., Dept. of Physics, 1 Brookings Dr., St. Louis, MO 63130

Mami Matsukawa, Cochair

Doshisha Univ., Lab. of Ultrasonic Electronics, Kyoto 610-0321, Japan

Chair's Introduction—7:30

Invited Papers

7:35

4aPA1. *In vivo* measurement of mass density and elasticity of cancellous bone using acoustic parameters for fast and slow waves. Takahiko Otani (Faculty of Eng., Doshisha Univ., Kyotanabe 610-0321 Japan)

Cancellous bone (spongy bone) is comprised of a porous network of numerous trabecular elements with soft tissue in the pore space. The effect of decreasing bone density, namely a symptom of osteoporosis, is greater for cancellous bone than for dense cortical bone (compact bone). Two longitudinal waves, the fast and slow waves, are clearly observed in cancellous bone, which correspond to ''waves of the first and second kinds'' as predicted by Biot's theory. According to experimental and theoretical studies, the propagation speed of the fast wave increases with bone density and that of the slow wave is almost constant. Experimental results show that the fast wave amplitude increases proportionally and the slow wave amplitude decreases inversely with bone density. However, the attenuation constant of the fast wave is almost independent of bone density and the attenuation constant of the slow wave increases with   bone   density.   The *in vivo* ultrasonic wave propagation path is composed of soft tissue, cortical bone, and cancellous bone and is modeled to specify the causality between ultrasonic wave parameters and bone mass density of cancellous bone. Then, mass density and elasticity are quantitatively formulated and estimated.

7:55

4aPA2. Is ultrasound appropriate to measure bone quality factors? Pascal Laugier (Univ. Pierre et Marie Curie., UMR CNRS 7623, 15 rue de l'Ecole de medecine, 7506 Paris, France, laugier@lip.bhdc.jussieu.fr)

Theoretical considerations support the concept that quantitative ultrasound variables measured in transmission are mainly determined by bone microstructure and material properties. All these properties are attributes of bone other than its bone mineral density (BMD) that may contribute to its quality and thus to strength or fragility. However, the limitations of this approach for a BMD-independent characterization of bone quality, long questioned, have become indisputable. Such considerations have prompted a research aiming at the development of new methods capable of measuring bone quality factors. New ultrasonic approaches are being investigated that use ultrasonic backscatter, guided waves, or nonlinear acoustics for studying bone microstructure or microdamage. These approaches, combined with sophisticated theoretical models or powerful computational tools, are advancing ideas regarding ultrasonic assessment of bone quality, which is not satisfactorily measured by x-ray techniques.

8:15

4aPA3. Scanning acoustic microscopy studies of cortical and trabecular bone in the femur and mandible. J. Lawrence Katz, Paulette Spence, Yong Wang (Univ. of Missouri-Kansas City, 650 E. 25th St., Kansas City, MO 64108, katzjl@umkc.edu), Anil Misra, Orestes Marangos (Univ. of Missouri-Kansas City, Kansas City, MO 64110), Dov Hazony (Case Western Reserve Univ., Cleveland, OH 44106), and Tsutomu Nomura (Niigata Univ. Grad. School of Medical and Dental Sci., Niigata, Japan)

Scanning acoustic microscopy (SAM) has been used to study the micromechanical properties of cortical and trabecular bone in both the human femur and mandible. SAM images vary in gray level reflecting the variations in reflectivity of the material under investigation. The reflection coefficient, $r = (Z2 - Z1)/(Z2 + Z1)$, where the acoustic impedance (AI), $Z = dv$, d is the materials local density and v is the speed of sound at the focal point; Z2 represents the AI of the material, Z1 that of the fluid coupling the acoustic wave from the lens to the material. Femoral cortical bone consists of haversian systems (secondary osteons) and interstitial lamellae, both of which show systematic variations of high and low AI from lamella to lamella. The lamellar components defining the edges of trabecular cortical bone exhibit the same lamellar variations as seen in cortical bone. Mandibular bone, while oriented perpendicular to the direction of gravitational attraction, exhibits the same cortical and trabecular structural organizations as found in the femur. It also exhibits the same systematic alternations in lamellar AI as found in femoral bone. Both femoral and mandibular cortical bone have transverse isotropic symmetry. Thus, modeling elastic properties requires only five independent measurements.

3241   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3241**

4aPA4. The interaction between ultrasound and human cancellous bone. Keith Wear (US Food and Drug Administration, 12720 Twinbrook Pkwy., Rockville, MD 20852, keith.wear@fda.hhs.gov)

Attenuation is much greater in cancellous bone than in soft tissues, and varies approximately linearly with frequency between 300 kHz and 1.7 MHz. At diagnostic frequencies (300 to 700 kHz), sound speed is slightly faster in cancellous bone than in soft tissues. A linear-systems model can account for errors in through-transmission-based measurements of group velocity due to frequency-dependent attenuation and dispersion. The dependence of phase velocity on porosity may be predicted from theory of propagation in fluid-filled porous solids. The dependence of phase velocity on frequency (negative dispersion) can be explained using a stratified two-component model. At diagnostic frequencies, scattering varies as frequency to the nth power where $3 > n > 3.5$. This may be explained by a model that represents trabeculae as finite-length cylindrical scatterers.

4aPA5. Dependence of phase velocity on porosity in cancellous bone: Application of the modified Biot-Attenborough model. Suk Wang Yoon and Kang Il Lee (Dept. of Phys. and Inst. of Basic Sci., SungKyunKwan Univ., Suwon 440-746, Republic of Korea)

This study aims to apply the modified Biot-Attenborough (MBA) model to predict the dependence of phase velocity on porosity in cancellous bone. The MBA model predicted that the phase velocity decreases nonlinearly with porosity. The optimum values for input parameters of the MBA model, such as compressional speed c_m of solid bone and phase velocity parameter s_2, were determined by comparing the prediction with the previously published measurements in human calcaneus and bovine cancellous bones. The value of the phase velocity parameter $s_2 = 1.23$ was obtained by curve fitting to the experimental data only for 53 human calcaneus samples with a compressional speed $c_m = 2500$ m/s of solid bone. The root-mean-square error (rmse) of the curve fit was 15.3 m/s. The optimized value of s_2 for all 75 cancellous bone samples (53 human and 22 bovine samples) was 1.42 with the rmse of 55 m/s. The latter fit was obtained by using $c_m = 3200$ m/s. Although the MBA model relies on empirical parameters determined from the experimental data, it is expected that the model can be usefully employed as a practical tool in the field of clinical ultrasonic bone assessment.

4aPA6. Simulation of fast and slow wave propagations through cancellous bone using three-dimensional elastic and Biot's trabecular models. Atsushi Hosokawa (Dept. of Elect. & Comp. Eng., Akashi Natl. Coll. of Tech., 679-3 Nishioka, Uozumi, Akashi, 674-8501 Hyogo, Japan, hosokawa@akashi.ac.jp)

The propagation of ultrasonic pulse waves in cancellous (trabecular) bone was numerically simulated by using three-dimensional finite-difference time-domain (FDTD) methods. In previous research [A. Hosokawa, J. Acoust. Soc. Am. **118**, 1782–1789 (2005)], two two-dimensional FDTD models, the commonly used elastic FDTD model and an FDTD model based on Biot's theory for elastic wave propagation in an isotropic fluid-saturated porous medium, were used to simulate the fast and slow longitudinal waves propagating through cancellous bone in the direction parallel to the main trabecular orientation. In the present study, the extended three-dimensional viscoelastic and Biot's anisotropic models were developed to investigate the effect of trabecular structure on the fast and slow wave propagations. Using the viscoelastic model of the trabecular frame comprised of numerous pore spaces in the solid bone, the effect of the trabecular irregularity, that is the scattering effect, on both the fast and slow waves could be investigated. The effect of the anisotropic viscous resistance of the fluid in the trabecular pore spaces on the slow wave could be considered using Biot's anisotropic model.

Contributed Papers

4aPA7. Ultrasonic characteristics of *in vitro* human cancellous bone. Isao Mano (OYO Electric Co., Ltd., Joyo 610-0101 Japan), Tadahito Yamamoto, Hiroshi Hagino, Ryota Teshima (Tottori Univ., Yonago 683-8503 Japan), Toshiyuki Tsujimoto (Horiba, Ltd., Kyoto 601-8510 Japan), and Takahiko Otani (Doshisha Univ., Kyotanabe 610-0321 Japan)

Cancellous bone is comprised of a connected network of trabeculae and is considered as an inhomogeneous and anisotropic acoustic medium. The fast and slow longitudinal waves are clearly observed when the ultrasonic wave propagates parallel to the direction of the trabeculae. The propagation speed of the fast wave increases with bone density and that of the slow wave is almost constant. The fast wave amplitude increases proportionally and the slow wave amplitude decreases inversely with bone density. Human *in vitro* femoral head was sectioned to 10-mm-thick slices perpendicularly to the femoral cervical axis. These cancellous bone samples were subjected to the ultrasonic measurement system LD-100 using a narrow focused beam. The propagation speed and the amplitude of the transmitted wave both for the fast and slow waves were measured at 1-mm intervals. The local bone density corresponding to the measured points was obtained using a microfocus x-ray CT system. Experimental results show that the propagation speeds and amplitudes for the fast and slow waves are characterized not only by the local bone density, but also by the local trabecular structure.

4aPA8. The effect of hydroxyapatite crystallite orientation on ultrasonic velocity in bovine cortical bone. Yu Yamato, Kaoru Yamazaki, Akira Nagano (Orthopaedic Surgery, Hamamatsu Univ. School of Medicine, 1-20-1 Hamamatsu Shizuoka 431-3192, Japan, yy14@hama-med.ac.jp), Hirofumi Mizukawa, Takahiko Yanagitani, and Mami Matsukawa (Doshisha Univ., Kyotanabe, Kyoto-fu 610-0321, Japan)

Cortical bone is recognized as a composite material of diverse elastic anisotropy, composed of hydroxyapatite (HAp) crystallite and type 1 collagen fiber. The aim of this study is to investigate the effects of HAp orientation on elastic anisotropy in bovine cortical bone. Eighty cubic samples were made from the cortical bone of two left bovine femurs. Longitudinal wave velocity in orthogonal three axes was measured using a conventional pulse echo system. For evaluating the orientation of HAp crystallite, x-ray diffraction profiles were obtained from three surfaces of each cubic sample. The preference of crystallites showed strong dependence on the direction of surface. The C-axis of crystallites was clearly preferred to the bone axial direction. The velocity in the axial direction was significantly correlated with the amounts of HAp crystallites aligning to the axial axis. The HAp orientation and velocity varied according to the microstructures of the samples. The samples with Haversian structure

showed larger preference of crystallites than plexiform samples in the axial direction. These results show clear effects of crystallites orientation on velocity.

4aPA9. Frequency variations of attenuation and velocity in cortical bone *in vitro*. Magali Sasso, Guillaume Haiat (Universite Paris 12, Laboratoire de Mecanique Physique, UMR CNRS 7052 B2OA, 61, avenue du General de Gaulle, 94010 Creteil, France), Yu Yamato (Hamamatsu Univ. School of Medicine, Hamamastu, Shizuoka, 431-3192, Japan), Salah Naili (Universite Paris 12, 94010 Creteil, France), and Mami Matsukawa (Doshisha Univ., Kyotanabe, Kyoto-fu 610-0321, Japan)

The development of ultrasonic characterization devices for cortical bone requires a better understanding of ultrasonic propagation in this heterogeneous medium. The aim of this work is to investigate the frequency dependence of the attenuation coefficient and of phase velocity and to relate them with bone microstructure and anatomical position. One hundred twenty parallelepipedic samples (4–11 mm side) have been cut from three bovine specimens and measured four times with repositioning in transmission with a pair of 8-MHz central frequency transducers. Phase velocity and BUA could be determined with acceptable precision: coefficients of variation of 0.8% and 13%, respectively. Velocity dispersion and BUA values are comprised between −13 and 40 m/s/MHz and 2 and 12 dB/MHz/cm, respectively. Negative dispersion values were measured (similarly to trabecular bone) for 2% of the measured samples. BUA values were found to be smaller in plexiform than in Haversian structure and higher for porotic structure. BUA values were found to be the greatest in the postero-lateral distal part and the smallest in the anterior-medial center part of the bone. The same tendency was found for velocity dispersion. Our results show the sensitivity of the frequency dependence of ultrasound to anatomical position and micro-architectural properties of bone.

4aPA10. 3-D numerical simulations of wave propagation in trabecular bone predicts existence of the Biot fast compressional wave. Guillaume Haiat (Universite Paris 12, Laboratoire de Mecanique Physique, UMR CNRS 7052 B2OA, 61, avenue du General de Gaulle, 94010 Creteil, France), Frederic Padilla, and Pascal Laugier (Universite Pierre et Marie Curie, 75006 Paris, France)

Trabecular bone is a poroelastic medium in which the propagation of two longitudinal waves (fast and slow) has been observed. The 3-D finite-difference time-domain simulations neglecting absorption coupled to 3-D microstructural models of 34 trabecular bone reconstructed from synchrotron radiation microtomography are shown to be suitable to predict both types of compressional wave in the three orthogonal directions. The influence of bone volume fraction (BV/TV) on the existence of the fast and slow waves was studied using a dedicated iterative image processing algorithm (dilation, erosion) in order to modify all 34 initial 3-D microstructures. An automatic criteria aiming at determining the existence of both wave modes was developed from the analysis of the transmitted signals in the frequency domain. For all samples, the fast wave disappears when bone volume fraction decreases. Both propagation modes were observed for BV/TV superior to a critical value for 2, 13, and 17 samples according to the direction of propagation. Above this critical value, the velocity of the fast (slow) wave increases (decreases) with BV/TV, consistent with Biot's theoretical predictions. This critical value decreases when the degree of anisotropy increases, showing coupling between structural anisotropy and the existence of the fast wave.

4aPA11. 3-D numerical simulation of wave propagation in porous media: Influence of the microstructure and of the material properties of trabecular bone. Guillaume Haiat (Universite Paris 12, Laboratoire de Mecanique Physique, UMR CNRS 7052 B2OA, 61, avenue du General de Gaulle, 94010 Creteil, France), Frederic Padilla, and Pascal Laugier (Universite Pierre et Marie Curie, 75006 Paris, France)

Finite-difference time domain simulations coupled to 3-D microstructural models of 30 trabecular bones reconstructed from synchrotron radiation microtomography were employed herein to compare and quantify the effects of bone volume fraction, microstructure, and material properties of trabecular bone on QUS parameters. Scenarios of trabecular thinning and thickening using an iterative dedicated algorithm allowed the estimation of the sensitivity of QUS parameters to bone volume fraction. The sensitivity to bone material properties was assessed by considering independent variations of density and stiffness. The effect of microstructure was qualitatively assessed by producing virtual bone specimens of identical bone volume fraction (13%). Both BUA and SOS show a strong correlation with BV/TV ($r^2 = 0.94$ $p10^{-4}$) and vary quasi-linearly with BV/TV at an approximate rate of 2 dB/cm/MHz and 11 m/s per % increase of BV/TV, respectively. Bone alterations caused by variation in BV/TV (BUA: 40 dB/cm.MHz; SOS: 200 m/s) is much more detrimental to QUS variables than that caused by alterations of material properties or diversity in microarchitecture (BUA: 7.8 dB/cm.MHz; SOS: 36 m/s). QUS variables are changed more dramatically by BV/TV than by changes in material properties or microstructural diversity. However, material properties and structure also appear to play a role.

4aPA12. Singular value decomposition-based wave extraction algorithm for ultrasonic characterization of cortical bone *in vivo*. Magali Sasso, Guillaume Haiat (Universite Paris 12, Laboratoire de Mecanique Physique, UMR CNRS 7052 B2OA, 61, avenue du General de Gaulle, 94010 Creteil, France), Maryline Talmant, Pascal Laugier (Universite Pierre et Marie Curie, 75006 Paris, France), and Salah Naili (Universite Paris 12, 94010 Creteil, France)

In the context of bone status assessment, the axial transmission technique allows ultrasonic evaluation of cortical bone using a multielement probe. The current processing uses the first arriving signal to evaluate the velocity while later contributions are potentially valuable and are not yet analyzed. However, all those contributions interfere, which disrupts the signal analysis. A novel ultrasonic wave extraction algorithm using a singular value decomposition method is proposed. This algorithm aims at characterizing a given energetic low-frequency (ELF) contribution observed *in vivo* at around 1 MHz. To evaluate the performances of the proposed algorithm, a simulated data set was constructed taking into account the influence of noise and of random interfering wavefront. The velocity of the ELF contribution is estimated on simulated datasets and compared to the input velocity. For a signal-to-noise ratio of 10 dB, the mean error associated with this method is 5.2%, to be compared with 34% with a classical signal analysis. The algorithm was also tested on real *in vivo* measurements. Results show the ability to accurately identify and possibly remove this wavefront contribution. Results are promising for multiple ultrasonic parameters evaluation from different wavefront contributions in our configuration.

4aPA13. Direct evaluation of cancellous bone porosity using ultrasound. Peiying Liu, Matthew Lewis, and Peter Antich (Grad. Program of Biomed. Eng., Univ. of Texas Southwestern Medical Ctr. at Dallas, 5323 Harry Hines Blvd, Dallas, TX 75390-9058)

Quantitative measurements of trabecular bone porosity would be a great advantage in diagnosis and prognosis of osteoporosis. This study focused on evaluation of the relationship between ultrasonic reflection and the density and porosity of cancellous bone. Theoretical simulation using MATLAB and Field II ultrasound simulator predicted a linear model of a material's porosity and parameters of the reflected ultrasonic signals. Ex-

perimentally, four plastic phantoms fabricated with different porosities were tested by a 5-MHz ultrasound transducer and the results agreed with simulations. Twelve specimens of bovine cancellous bone were measured. The porosities of these specimens were estimated by calculating the ratio of the mass in air to the wetted mass when they were immersed in water and all the air was removed from the pores. Among all the parameters, the peak value of the reflected ultrasound signal demonstrated highly significant linear correlations with porosity ($R^2 = 0.911$) and density ($R^2 = 0.866$). This encouraging result shows that this technique has the potential to be used to monitor porosity changes noninvasively for clinical purpose such as noninvasive assessment of osteoporotic fracture risk. [Work supported by Pak Foundation.]

11:35

4aPA14. Transmission of ultrasound through bottlenose dolphin (*tursiops truncatus*) **jaw and skull bone.** Michael D. Gray, James S. Martin, and Peter H. Rogers (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, michael.gray@gtri.gatech.edu)

Measurements of ultrasound transmission through jaw (pan bone) and skull (temporal fossa) samples from an Atlantic bottlenose dolphin were performed as part of an investigation of the feasibility of performing *in vivo* elastography on cetacean head tissues. The pan bone and temporal fossa are both relatively thin and smooth, and are therefore of interest for acoustic imaging of the underlying tissues of the ear and brain, respectively. Field scan data will be presented, showing the influence of the bone on the quality of the focus and overall pressure levels generated by a spherically focused single-element transducer. [Work supported by ONR.]

11:50

4aPA15. Assessment of bone health by analyzing propagation parameters of various modes of acoustic waves. Armen Sarvazyan, Vladimir Egorov, and Alexey Tatarinov (Artann Labs., Inc., 1457 Lower Ferry Rd., West Trenton, NJ 08618)

A method for assessment of bone based on comprehensive analysis of waveforms of ultrasound signals propagating in the bone is presented. A set of ultrasound propagation parameters, which are differentially sensitive to bone material properties, structure, and cortical thickness, are evaluated. The parameters include various features of different modes of acoustics

waves, such as bulk, surface, and guided ultrasonic waves in a wide range of carrier frequencies. Data processing algorithms were developed for obtaining axial profiles of waveform parameters. Such profiles are capable of revealing axial heterogeneity of long bones and spatially nonuniform pathological processes, such as osteoporosis. The examination procedure may include either one long bone in the skeleton, like the tibia, radius of the forearm, etc., or several bones in sequence to provide a more comprehensive assessment of the skeletal system. Specifically, for tibia assessment, a multi-parametric linear classifier based on a DEXA evaluation of skeleton conditions has been developed. Preliminary results of the pilot clinical studies involving 149 patients have demonstrated early stages of osteoporosis detection sensitivity of 80% and specificity of 67% based on DEXA data as the gold standard. [Work was supported by NIH and NASA grants.]

12:05

4aPA16. Toward bone quality assessment: Interference of fast and slow wave modes with positive dispersion can account for the apparent negative dispersion. James G. Miller, Karen Marutyan, and Mark R. Holland (Dept. of Phys., Washington Univ., St. Louis, MO 63130)

The goal of this work was to show that apparently negative dispersion in bone can arise from interference between fast wave and slow wave longitudinal modes, each of positive dispersion. Simulations were carried out using two approaches, one based on the Biot-Johnson model and the second independent of that model. The resulting propagating fast wave and slow wave modes accumulated phase and suffered loss with distance traveled. Results of both types of simulation served as the input into a phase and magnitude spectroscopy algorithm (previously validated with experimental data) in order to determine the phase velocity as a function of frequency. Results based on both methods of simulation were mutually consistent. Depending upon the relative magnitudes and time relationships between the fast and slow wave modes, the apparent phase velocities as functions of frequency demonstrated either negative or positive dispersions. These results appear to account for measurements from many laboratories that report that the phase velocity of ultrasonic waves propagating in cancellous bone decreases with increasing frequency (negative dispersion) in about 90% of specimens but increases with frequency in about 10%. [Work supported in part by Grant NIH R37HL40302.]

Session 4aPP

Psychological and Physiological Acoustics and ASA Committee on Standards: New Insights on Loudness and Hearing Thresholds

Rhona P. Hellman, Cochair

Northeastern Univ., Dept. of Speech-Language Pathology and Audiology, Boston, MA 02115

Yôiti Suzuki, Cochair

Tohoku Univ., Research Inst. of Electrical Communication, Katarahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan

Chair's Introduction—8:30

Invited Papers

8:35

4aPP1. Threshold of hearing for pure tones between 16 and 30 kHz. Kaoru Ashihara (AIST, Tsukuba Central, 6 Tsukuba, Ibaraki 305-8566 Japan, ashihara-k@aist.go.jp)

Hearing thresholds for pure tones were obtained at 2-kHz intervals between 16 and 30 kHz in an anechoic chamber. Measured 50 cm from the sound source, the maximum presentation sound pressure level ranged from 105 to 112 dB depending on the frequency. To prevent the listener from detecting the quantization noise or subharmonic distortions at low frequencies, a pink noise was added as a masker. Using a 3-down 1-up transformed up-down method, thresholds were obtained at 26 kHz for 10 out of 32 ears. Even at 28 kHz threshold values were obtained for 3 ears, but none were observed for a tone at 30 kHz. Above 24 kHz, the thresholds always exceeded 90 dB SPL. Between 16 and 20 kHz thresholds increased abruptly, whereas above 20 kHz the threshold increase was more gradual.

8:55

4aPP2. Do we have better hearing sensitivity than people in the past? Kenji Kurakata and Tazu Mizunami (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki, 305-8566 Japan, kurakata-k@aist.go.jp)

Our hearing sensitivity to tones of various frequencies declines progressively as we become older. ISO 7029 describes a method for calculating expected hearing threshold values as a function of age. However, more than 30 years have passed since the ISO standard source data were published. An earlier paper of the present authors [K. Kurakata and T. Mizunami, Acoust. Sci. Technol. **26**, 381–383 (2005)] compared hearing threshold data of Japanese people in recent years to the ISO standard values. The results of that comparison showed that the age-related sensitivity decrease of Japanese people was smaller, on average, than those described in the standard. A large discrepancy was apparent at 4000 and 8000 Hz: more than 10 dB for older males. In response to that inference, the ISO/TC43/WG1 "threshold of hearing" initiated a project in 2005 to explore the possibility of updating ISO 7029. This paper presents a summary of those comparison results of audiometric data and the work in WG1 for revising the standard.

9:15

4aPP3. Reliability and frequency specificity of auditory steady-state response. Masaru Aoyagi, Tomoo Watanabe, Tsukasa Ito, and Yasuhiro Abe (Dept. of Otolaryngol., Yamagata Univ. School of Medicine, 2-2-2 Iida-Nishi, Yamagata, 990-9585, Japan)

The reliability and frequency specificity of 80-Hz auditory steady-state response (80-Hz ASSR) elicited by sinusoidally amplitude-modulated (SAM tones) tones and detected by phase coherence were evaluated as a measure of the hearing level in young children. The 80-Hz ASSR at a carrier frequency of 1000 Hz was monitored in 169 ears of 125 hearing-impaired children and auditory brainstem response (ABR) elicited by tone pips was examined in 93 ears. Both responses were examined during sleep, and the thresholds were compared with the behavioral hearing threshold, which was determined by standard pure-tone audiometry or play audiometry. In 24 ears with various patterns of audiogram, 80-Hz ASSRs were examined at different carrier frequencies, and the threshold patterns were compared with the audiograms to investigate the frequency specificity of ASSR. The correlation coefficient between the threshold of 80-Hz ASSR and pure-tone threshold ($r = 0.863$) was higher than that for ABR ($r = 0.828$). The threshold patterns of 80-Hz ASSR clearly followed the corresponding audiogram patterns in all types of hearing impairment. These findings suggest that 80-Hz ASSR elicited by SAM tones and detected by phase coherence is a useful audiometric device for the determination of hearing level in a frequency-specific manner in children.

4a FRI. AM

3245   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3245**

9:35

4aPP4. Use of perceptual weights to test a model of loudness summation. Lori J. Leibold and Walt Jesteadt (Boys Town Natl. Res. Hospital, 555 N 30th St., Omaha, NE 68131)

We recently examined the contribution of individual components of a multitone complex to judgments of overall loudness by computing the perceptual weight listeners assign to each component in a loudness-matching task [Leibold *et al.*, J. Acoust. Soc. Am. **117**, 2597 (2005)]. Stimuli were five-tone complexes centered on 1000 Hz, with six different logarithmic frequency spacings, ranging from 1.012 to 1.586. When all components fell within the same critical band, weights varied little across components. In contrast, the range of weights increased with increasing frequency separation, with greater weight given to the lowest and highest frequency components. Perceptual weights were largely in agreement with the Moore *et al.* loudness model [J. Audio Eng. Soc. **45**, 224–237 (1997)], except at the widest bandwidth. In the current study we further examined predictions of the loudness model, focusing on the widest frequency-spacing condition. Masked thresholds and jnds for intensity discrimination were measured for each component and were compared to weights. The model predicts more interaction in the widely spaced conditions than simple critical band models, but underestimates the true interactions in conditions where components are widely spaced. Central factors appear to influence loudness, masked thresholds, and intensity discrimination in these conditions. [Work supported by NIH/NIDCD.]

9:55

4aPP5. Increased loudness effect at the absolute threshold of hearing. Junji Yoshida, Masao Kasuga (Grad. School of Eng., Utsunomiya Univ., 7-1-2 Yoto, Utsunomiya-shi, Tochigi-ken, 321-8585, Japan), and Hiroshi Hasegawa (Utsunomiya Univ., Tochigi-ken, 321-8585, Japan)

This study investigated the effects of a previous sound on loudness at the absolute threshold of hearing. Change of the absolute threshold of hearing was measured when a pure tone preceded the test tone in the measurement of the threshold. The previous sound at 60 dB SPL was presented first in one ear, followed by the presentation of the test sound in either the contralateral or ipsilateral ear at an interval of 0.5 s. Both the previous and test sounds had the same frequency of 500 Hz, and the same duration of 3 s. The change of the threshold was obtained from the difference between the thresholds with and without the previous sound. The threshold was found to be decreased significantly by approximately 2 dB when the previous sound was presented in the contralateral ear. On the other hand, the threshold was only slightly changed when the previous sound was presented in the ipsilateral ear. These experiment results suggest that loudness was increased by perceiving the previous sound in the contralateral ear.

10:15−10:25   Break

10:25

4aPP6. Induced loudness reduction. Michael Epstein (Auditory Modeling and Processing Lab., Inst. for Hearing, Speech & Lang., Dept. of Speech-Lang. Path. and Aud., Northeastern Univ., Boston, MA 02115) and Mary Florentine (Northeastern Univ., Boston, MA)

Induced loudness reduction (ILR) is a phenomenon by which a preceding higher-level tone (an inducer tone) reduces the loudness of a lower-level tone (a test tone). Under certain conditions, ILR can result in loudness reductions of 10 to 15 phons for pure tones. The strength of this effect depends on a number of parameters including: (1) the levels of both the inducer and test tones; (2) the frequency separation between the inducer and test tones; (3) the durations of the inducer and test tones; (4) the time separation between the inducer and test tones; (5) individual differences; and, possibly (6) the number of exposures to inducers. Because of the sensitivity of ILR to procedural conditions, it is quite important to consider the potential effects of ILR when considering any experimental design in which level varies. The understanding of ILR has also given insight into a number of unexplained discrepancies between data sets that were collected using different procedures. In fact, some of the variability known to affect loudness judgments may be due to ILR. [Work supported by NIH/NIDCD Grant R01DC02241.]

10:45

4aPP7. Loudness growth in individual listeners with hearing loss. Jeremy Marozeau and Mary Florentine (Commun. Res. Lab., Inst. for Hearing, Speech & Lang., Northeastern Univ., Boston, MA 02115)

Recent research indicates that there are large individual differences in how loudness grows with level for listeners with sensorineural hearing losses of primarily cochlear origin. Studies of loudness discomfort level suggest that loudness for most of these impaired listeners approaches the loudness experienced by the normal listeners at high levels. Loudness growth at low levels is still controversial. Although it is now well established that loudness at threshold is greater than zero, its exact value is unknown. If this value is the same for normal and impaired listeners, then the loudness growth for the impaired listeners must be steeper in order to catch up to normal at high levels. This phenomenon is called recruitment. If the loudness at threshold for impaired listeners is higher than that for normal listeners, then the impaired listeners will no longer be able to perceive sounds as soft. This phenomenon is called softness imperception. Results from two experiments suggest that: (1) individual differences are more important for impaired listeners than for normal listeners; (2) some impaired listeners seem to show recruitment, others softness imperception; and (3) averaging the results across the impaired listeners will mask these differences. [Work supported by NIH/NIDCD grant R01DC02241.]

4aPP8. Growth of loudness in cochlear implant listening. Robert L. Smith and Nicole Sanpetrino (Inst. for Sensory Res. and Dept. of Biomed. and Chemical Eng., Syracuse Univ., 621 Skytop Rd., Syracuse NY, 13244)

Cochlear implants (CIs) roughly mimic the transformation from sound frequency to cochlear place that occurs in acoustic hearing. However, CIs are relatively less capable of creating the intensive transformations that normal peripheral auditory processing provides. This is partly because CIs have small operating ranges on the order of 10:1 in electric current, compared to the 1 000 000:1 operating range for sound-pressure level (SPL) in acoustic hearing. Furthermore, loudness in acoustic hearing grows as a compressive power function of SPL. In contrast, loudness reportedly grows as a more expansive function of current for CI users, i.e., a power law with a large exponent or an exponential function. Our results, obtained using the minimally biased method of magnitude balance without an arbitrary standard, reveal a previously unreported range of shapes of CI loudness functions, going from linear to power laws with exponents of 5 or more. The shapes seem related in part to the etiology of deafness preceding cochlear implantation, although the shapes can vary with stimulating conditions within a subject. Furthermore, differential sensitivity to changes in current appears to be related to the shape of the corresponding loudness function. Implications for sound processing in electric and acoustic hearing will be discussed.

FRIDAY MORNING, 1 DECEMBER 2006

MOLOKAI ROOM, 8:00 A.M. TO 12:00 NOON

Session 4aSC

Speech Communication: Perception (Poster Session)

Katsura Aoyama, Cochair
Texas Tech. Univ., Health Science Ctr., Dept. of Speech-Language and Hearing Science, Lubbock, TX 79430-6073

Masato Akagi, Cochair
JAIST, School of Information Science, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon.

4aSC1. Functional load of segments and features. Mafuyu Kitahara (School of Law, Waseda Univ., 1-6-1 Nishiwaseda, Shinjuku-Ku, Tokyo, Japan)

The present paper proposes a new measure of functional load for segments and features. In a nut shell, it is based on word frequencies and the number of minimal pairs in which the relevant segment/feature is crucial in distinction. For example, minimal pairs distinguished only by /t/ are most frequent in English while those distinguished by /k/ are most frequent in Japanese. As for functional load of features, single-feature contrasts and multiple-feature contrasts are incorporated in the calculation. In Japanese, [high] alone distinguishes the largest number of minimal pairs while [voice] distinguishes words most frequently in cooperation with other features. Word frequency and familiarity database for English and Japanese are used to observe the commonalities and differences in both languages with respect to the proposed measure of functional load. This line of analysis suggests a better account for a certain phonological process being more typical in one language but not in the other. Functional load can be thought of as a part of the top-down information from the lexicon, which interacts with the bottom-up perceptual information in the process of spoken word recognition. Not only the ease of articulation and perceptual salience but also the functional load drives phonological processes.

4aSC2. Cortical representation of processing Japanese phonemic and phonetic contrasts. Seiya Funatsu (The Prefectural Univ. of Hiroshima, 1-1-71 Ujinahigashi, Minami-ku, Hiroshima, 734-8558 Japan), Satoshi Imaizumi (The Prefectural Univ. of Hiroshima, Gakuen-machi Mihara, 723-0053 Japan), Akira Hashizume, and Kaoru Kurisu (Hiroshima Univ., Minami-ku Hiroshima, 734-8551 Japan)

This study investigated how Japanese speakers process phonemic and phonetic contrasts using voiced and devoiced vowel /u/ and /÷u/. During six oddball experiments, brain responses were measured using magnetoencephalography. Under the phonemic condition, a frequent stimulus /ch÷ita/ was contrasted with a deviant /ts÷uta/, and a frequent /ts÷uta/ with a deviant /ch÷ita/. Under the phonetic condition, a frequent /ts÷uta/ was contrasted with a deviant /tsuta/, and a frequent /tsuta/ with a deviant /ts÷uta/. Under the segment condition, vowel segments, /÷u/ and /u/, extracted from spoken words, were contrasted. The subjects were 13 native Japanese speakers. The equivalent current dipole moment (ECDM) was estimated from the mismatch field. Under the phonetic condition, the ECDM elicited by the voiced deviant was significantly larger than that elicited by the devoiced deviant in both hemispheres ($p<0.01$), while there were no significant deviant-related differences in ECDM under the phonemic condition in both hemispheres. Under the segment condition, the ECDM elicited by the voiced deviant and devoiced deviant did not differ significantly in either hemispheres. These results suggested that the

4a FRI. AM

ECDM asymmetries between the voiced and the devoiced deviant observed under the phonetic condition did not originate from the acoustical difference itself, but from the phonetic environment.

4aSC3. Evaluating a model to estimate breathiness in vowels.

Rahul Shrivastav, Arturo Camacho, and Sona Patel (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611)

The perception of breathiness in vowels is cued by changes in aspiration noise (AH) and the open quotient (OQ) [Klatt and Klatt, J. Acoust. Soc. Am. **87**(2), 820–857 (1990)]. A loudness model can be used to determine the extent to which AH masks the harmonic components in voice. The resulting partial loudness (PL) and loudness of AH (noise loudness; NL) have been shown to be good predictors of perceived breathiness [Shrivastav and Sapienza, J. Acoust. Soc. Am. **114**(1), 2218–2224 (2005)]. The levels of AH and OQ were systematically manipulated for ten synthetic vowels. Perceptual judgments of breathiness were obtained and regression functions to predict breathiness from NL/PL were derived. Results show breathiness to be a power function of NL/PL when NL/PL is above a certain threshold. This threshold appears to be affected by the stimulus pitch. A second experiment was conducted to determine if the resulting power function could be used to estimate breathiness in natural voices. The breathiness of novel stimuli, both natural and synthetic, was determined in a listening test. For comparison, breathiness for the same stimuli was also estimated using the power function obtained previously. Results show the extent to which findings can be generalized. [Research supported by NIH/R21DC006690.]

4aSC4. Predicting vowel inventories: The dispersion-focalization theory revisited.

Roy Becker (Dept. of Linguist., Univ. of California Los Angeles, 3125 Campbell Hall, Los Angeles, CA 90095-1543)

A revision of the dispersion-focalization theory (DFT) [Schwartz *et al.*, J. Phonetics **25**, 233–253 (1997)] is presented. Like DFT, the current computational model incorporates the center of gravity effect (COG) of 3.5-Bark spectral integration, but it deviates from DFT in that the COG contributes to the actual values and reliability weights of the perceived formants of vowels. The COG is reinterpreted as a domain of acceleration towards formant merger: the percepts of formants less than 3.5 Barks apart are perturbed towards one another in a nonlinear yet continuous fashion and their weights are increased, but perceptual merger and weight maximization occur only when the acoustic distance is about 2 Bark. Like other dispersion-based models, inventories are evaluated predominantly according to the least dispersed vowel pair, where dispersion is measured as the weighted Euclidean distance between the vowels coordinates (the first two perceived formants). Yet in the current model the weights are determined dynamically, in a well-principled manner. This model improves existing models in predicting certain universal traits, such as series of front rounded vowels in large vowel inventories, as emergent properties of certain local maxima of the inventory dispersion evaluation function, without sacrificing predictive adequacy for smaller inventories.

4aSC5. Matching fundamental and formant frequencies in vowels.

Peter F. Assmann (School of Behavioral and Brain Sci., Univ. of Texas at Dallas, Box 830688, Richardson, TX 75083), Terrance M. Nearey (Univ. of AB, Edmonton, AB, Canada T6E 2G2), and Derrick Chen (Univ. of Texas at Dallas, Richardson, TX 75083)

In natural speech, there is a moderate correlation between fundamental frequency (F0) and formant frequencies (FF) associated with differences in larynx and vocal tract size across talkers. This study asks whether listeners prefer combinations of mean F0 and mean FF that mirror the covariation of these properties. The stimuli were vowel triplets (/i/-/a/-/u/) spoken by two men and two women and subsequently processed by Kawahara's STRAIGHT vocoder. Experiment 1 included two continua, each containing 25 vowel triplets: one with the spectrum envelope (FF) scale factor fixed at 1.0 (i.e., unmodified) and F0 varied over ±2 oct, the other with F0 scale factor fixed at 1.0 and FF scale factors between 0.63 and 1.58. Listeners used a method of adjustment procedure to find the "best voice" in each set. For each continuum, best matches followed a unimodal distribution centered on the mean F0 or mean FF (F1, F2, F3) observed in measurements of vowels spoken by adult males and females. Experiment 2 showed comparable results when male vowels were scaled to the female range and vice versa. Overall the results suggest that listeners have an implicit awareness of the natural covariation of F0 and FF in human voices.

4aSC6. Acoustic cues for distinguishing consonant sequences in Russian.

Lisa Davidson and Kevin Roon (Linguist. Dept., New York Univ., 719 Broadway, 4th Fl, New York, NY 10003, lisa.davidson@nyu.edu)

In Russian, the same consonant sequences are permitted in various environments. Consequently, the presence of a word boundary or reduced vowel can be phonologically contrastive (e.g. [zədɐvatʲ] "to assign," [zdɐvatʲ] "to turn in"), and both learners and experienced listeners likely rely on fine acoustic cues to discriminate the phonotactic structures they hear. In this study, the acoustic characteristics of consonant sequences are examined to establish which cues may distinguish (a) word-initial clusters (C1C2); (b) consonant-schwa-consonant sequences (C1VC2); and (c) sequences divided by a word boundary (C1#C2). For all three sequence types, native Russian speakers produced phrases containing three categories of target sequences: stop+consonant, fricative+consonant, and nasal +consonant. Results show no significant differences in C1 burst duration for initial stops, though a longer interconsonantal duration is a reliable cue to schwa presence in C1VC2. C2 is significantly longer in C#C than in other sequences. For clusters, when C1 is a stop, there are no significant differences in duration with other sequence types, but other C1's are significantly shorter. This suggests that articulatory overlap, which may lead to C1 shortening for fricative or nasal-initial clusters, is reduced in stop-initial clusters to ensure that the stop is released and recoverable. [Research supported by NSF.]

4aSC7. Lipread me now, hear me better later: Crossmodal transfer of talker familiarity effects.

Kauyumari Sanchez, Lawrence D. Rosenblum, and Rachel M. Miller (Dept. of Psych., Univ. of California, Riverside, Riverside, CA)

There is evidence that for both auditory and visual speech perception (lipreading) familiarity with a talker facilitates speech recognition [Nygaard *et al.*, Psychol. Sci. **5**, 42 (1994); Yakel *et al.*, Percept. Psychophys. **62**, 1405 (2000)]. Explanations of these effects have concentrated on the retention of talker information specific to each of these modalities. It could be, however, that some amodal, talker-specific articulatory style information is retained to facilitate speech perception in both modalities. If this is true, then experience with a talker in one modality should facilitate perception of speech from that talker in the other modality. To test this prediction, subjects were given one hour of experience lipreading from a talker and were then asked to recover speech-in-noise from either this same talker or from a different talker. Results revealed that subjects who lipread and heard speech from the same talker performed better on the speech-in-noise task than did subjects who lipread from one talker and then heard speech from a different talker.

4aSC8. Acoustic patterns of Japanese voiced velar stops.

James Dembowski and Katsura Aoyama (Dept. Speech-Lang. & Hearing Sci., Texas Tech Univ. Health Sci. Ctr., 3601 4th St., Lubbock, TX 79430-6073, james.dembowski@ttuhsc.edu)

This presentation focuses on Japanese voiced velar /g/. The phoneme /g/ in VCV contexts is said to be characterized by a distinctive wedge-shaped formant pattern in which F2 and F3 converge toward one fre-

quency in the transition from vowel to stop closure, and then diverge as the vocal tract opens from the stop release to the following vowel. This pattern was examined using acoustic and kinematic data from an x-ray microbeam database of Japanese speakers, which is comparable to the English language x-ray microbeam speech production database [Hashi *et al.*, J. Acoust. Soc. Am. **104**, 2426–2437 (1998)]. Japanese speakers produced the expected wedge-shaped formant pattern in isolated VCV nonsense syllables, but rarely, if ever, produced this pattern in connected speech. In contrast, English speakers more frequently produced the expected formant pattern in connected speech, though the pattern was less reliably present than in isolated VCVs and varied considerably within and across speakers. These observations highlight substantial differences between controlled laboratory speech and meaningful connected speech, as well as differences in the ways that phonemes are manifested by different linguistic communities. These data also potentially illuminate the relationship among phonetic, acoustic, and kinematic levels of speech production.

4aSC9. Finding perceptual categories in multidimensional acoustic spaces.
Eric Oglesbee and Kenneth de Jong (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405, eoglesbe@indiana.edu)

Examining phonetic categorization in multidimensional acoustic spaces poses a number of practical problems. The traditional method of forced identification of an entire stimulus space becomes prohibitive when the number and size of acoustic dimensions becomes increasingly large. In response to this, Iverson and Evans [ICPhS (2003)] proposed an adaptive tracking algorithm for finding best exemplars of vowels in a multidimensional acoustic space. Their algorithm converged on best exemplars in a relatively small number of trials; however, the search method took advantage of special properties of the vowel space in order to achieve rapid convergence. In this paper, a more general multidimensional search algorithm is proposed and analyzed for inherent biases. Then, using the algorithm, the phonetic categorization of /p/ and /b/ in a five-dimensional acoustic space by native speakers of English is tested. Results showed that (a) there were no substantial long-term biases in the search method and (b) the algorithm appeared to identify important acoustic dimensions in the identification of /p/ and /b/ using relatively few trials. [Work supported by NSF BCS-04406540.]

4aSC10. On the perception of epenthetic stops in American English.
Amalia Arvaniti, Ryan Shosted, and Cynthia Kilpatrick (Dept. of Linguist., UCSD, 9500 Gilman Dr., La Jolla, CA 92093, amalia@ling.ucsd.edu)

This study examines the perception of epenthetic stops in American English. Stimuli were created from *prince*, *prints*, *mince*, *mints*, *quince*, and *quints* by removing all traces of /t/ and splicing in 0–72 ms of silence, in 12-ms steps, with or without a following burst. Subjects saw the minimal pairs on screen and selected the word they heard. It was hypothesized that stimuli with bursts and longer closure would result in more *t* responses (*prince* identified as *prints*) and that frequent words (*prince/prints*) would be more difficult to distinguish than infrequent words (*quince/quints*), as our production results suggest that frequent pairs are more likely to be produced similarly. Results from 19 subjects show shorter response times with longer silence intervals, but no effect of burst or stimulus identity. Similarly, stimuli with bursts were not identified as *nts* words more frequently than those without. Generally, stimuli originating from *nts* words were more likely to be identified as such if they contained a burst, while the opposite was true for stimuli from *nce* words.

Finally, frequent words were less likely to be correctly identified than infrequent words, suggesting that /t/ epenthesis is not as widespread throughout the lexicon as generally believed.

4aSC11. Phonetic alignment to visual speech.
Rachel M. Miller, Lawrence D. Rosenblum, and Kauyumari Sanchez (Dept. of Psych., Univ. of California, Riverside, Riverside, CA 92521)

Talkers are known to produce allophonic variation based, in part, on the speech of the person with whom they are talking. This subtle imitation, or phonetic alignment, occurs during live conversation and when a talker is asked to shadow recorded words [e.g., Shockley, *et al.*, Percept. Psychophys. **66**, 422 (2004)]. What is yet to be determined is the nature of the information to which talkers align. To examine whether this information is restricted to the acoustic modality, experiments were conducted to test if talkers align to visual speech (lipread) information. Normal-hearing subjects were asked to watch an actor silently utter words, and to identify these words by saying them out loud as quickly as possible. These shadowed responses were audio recorded and naive raters compared these responses to the actors auditory words (which had been recorded along with the actors visual tokens). Raters judged the shadowed words as sounding more like the actors words than did baseline words, which had been spoken by subjects before the shadowing task. These results show that phonetic alignment can be based on visual speech, suggesting that its informational basis is not restricted to the acoustic signal.

4aSC12. New anthropomorphic talking robot—investigation of the three-dimensional articulation mechanism and improvement of the pitch range.
Kotaro Fukui, Yuma Ishikawa, Takashi Sawa, Eiji Shintaku (Dept. of Mech. Eng., Waseda Univ., 3-4-1 Ookubo, Shinjuku-ku, Tokyo, Japan), Masaaki Honda (Waseda Univ., Saitama, Japan), and Atsuo Takanishi (Waseda Univ., Shinjuku-ku, Tokyo, Japan)

We developed a new three-dimensional talking robot WT-6 (Waseda Talker-No. 6), which generates speech sounds by mechanically simulating articulatory motions as well as aeroacoustic phenomenon in the vocal tract. WT-6 consists of a 17-DOF mechanism (1-DOF lungs, 5-DOF vocal cords, 1-DOF jaws, 5-DOF tongue, and 4-DOF lips). It has 3-D lips, tongue, jaw, and velum, which form the 3-D vocal tract structure. It also has an independent jaw opening/closing mechanism, which controls the relative tongue position in the vocal tract as well as the oral opening. The previous robot, which had a 2-D tongue [J. Acoust. Soc. Am. **117**, 2541 (2005)], was not enough to realize precise closure to produce humanlike consonants such as /s/ or /r/. The new tongue, which could be controlled to form the 3-D shape, makes it possible to produce more realistic vocal tract shape. The vocal cord model was also improved by adding a new pitch control mechanism pushing from the side of the vocal cords. The pitch range becomes broader than the previous robot, which is enough to reproduce normal human speech. Preliminary experimental results showed that synthesized speech quality improves for vowels /a/, /u/ and /o/. Some experimental results and video demonstration of the talking robot will be presented.

4aSC13. The role of familiarity, semantic context, and amplitude-modulation on sentence intelligibility.
Tom Carrell (Univ. of Nebraska—Lincoln, Lincoln, NE 68583, tcarrell@unl.edu) and Dawna Lewis (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

Amplitude modulation has been demonstrated to greatly improve the intelligibility of time-varying sinusoidal (TVS) sentences [Carrell and Opie, Percept. Psychophys. **52** (1992); Barker and Cooke, Speech Commun. **27** (1999); Hu and Wang, Proceedings of ICASSP-02 (2002)]. It has been argued that the improvement is due to a bottom-up process that causes the acoustically independent components of the sentences to be perceptually grouped for further analysis by the auditory system. A recent study [Shapley and Carrell, J. Acoust. Soc. Am. **118** (2005)] indicated that

semantic information did not influence intelligibility levels of TVS or modulated TVS sentences. In virtually all other studies in which speech was distorted or degraded its intelligibility was improved by appropriate semantic context [Miller, *et al.*, JEP **41** (1951)]. It is possible that listeners' unfamiliarity with TVS speech might account for the difference. With one exception every study that has employed this type of stimulus provided the listener with very few practice sentences [Lewis, AAA (2005)]. The present experiment manipulated listeners' familiarity with TVS sentences to test this notion. Participants were presented with high- and low-predictability TVS and modulated-TVS sentences. Familiarity had a large effect on perception and intelligibility. Interpretation of previous findings is reconsidered in this light.

4aSC14. On the relation of apparent naturalness to phonetic perceptual resolution of consonant manner. Robert E. Remez, Claire A. Landau, Daria F. Ferro, Judith Meer, and Kathryn Dubowski (Dept. of Psych., Barnard College, 3009 Broadway, New York, NY 10027)

How does the qualitative experience of speech influence phonetic perception? Our perceptual studies of consonant place and voicing have revealed a dichotomous relation between phonetic sensitivity and naturalness. Auditory quality and phonetic sensitivity sometimes co-vary, while in other conditions phonetic sensitivity is indifferent to huge variation in naturalness. New tests are reported extending the research to the dimension of manner, a contrast correlated with qualitatively distinct acoustic constituents in normal production. Speech synthesis techniques were used to create naturalness variants through (1) variation in the excitation of a synthetic voicing source and (2) variation in the bandwidth of the formant centers. Listeners calibrated the relative naturalness of items drawn from the test series, and the acuity of perceivers to the contrast between fricative and stop manner was estimated with cumulative d' across the series in identification tests. Combined with our prior findings, these new tests show how intelligibility and naturalness can be either perceptually orthogonal or contingent aspects of consonant dimensions, offering a tool to understand normative functions in speech perception. [Research supported by NIH (DC00308).]

4aSC15. Effects of signal levels on vowel formant discrimination for normal-hearing listeners. Chang Liu (Dept. of Commun. Sci. and Disord., Wichita State Univ., Wichita, KS 67260)

The goal of this study was to examine effects of signal levels on vowel formant discrimination. Thresholds of formant discrimination were measured for F1 and F2 of four vowels in isolated vowels and sentences at three intensity levels: 70, 85, and 100 dB SPL for normal-hearing listeners using a three-interval, two-alternative forced-choice procedure with a two-down, one-up tracking algorithm. Results showed that formant thresholds were significantly affected by formant frequency, linguistic context, and signal levels. Thresholds of formant discrimination were increased as formant frequency increased and linguistic context changed from isolated vowels to sentences. The signal level indicated a rollover effect in which formant thresholds at 85 dB SPL are lower than at either 70 or 100 dB SPL in both isolated vowels and sentences. This rollover level effect could be due to reduced frequency selectivity and reduction in active cochlear nonlinearity at high signal levels for normal-hearing listeners. Excitation and loudness models will be explored to account for the level effect on formant discrimination of isolated vowels.

4aSC16. Study on nonaudible murmur speaker verification using multiple session data. Mariko Kojima, Hiromichi Kawanami, Hiroshi Saruwatari, Kiyohiro Shikano (Nara Inst. of Sci. and Technol. 8916-5 Takayama-cho Ikoma-shi, Nara, 630-0192 Japan), and Tomoko Matsui (The Inst. of Statistical Mathematics, Minato-ku, Tokyo, 106-8569 Japan)

A study on speaker verification with nonaudible murmur (NAM) segments using multiple session data was conducted. NAM is different from normal speech and is difficult for other people to catch. Therefore, a text-dependent verification strategy can be used in which each user utters her/

his own keyword phrase so that not only speaker-specific but also keyword-specific acoustic information is utilized. A special device called a NAM microphone worn on the surface of the skin below the mastoid bone is used to catch NAM because it is too low to be recorded using ordinary microphones. However, it is tolerant to exterior noise. This strategy is expected to yield relatively high performance. NAM segments, which consist of multiple short-term feature vectors, are used as input vectors to capture keyword-specific acoustic information well. To handle segments with a large number of dimensions, a support vector machine (SVM) is used. In experiments using NAM data uttered by 19 male and 10 female speakers in several different sessions, robustness against session-to-session data variation is examined. The effect of segment length is also investigated. The proposed approach achieves equal error rates of 0.04% (male) and 1.1% (female) when using 145-ms-long NAM segments.

4aSC17. Sequential contrast or compensation for coarticulation? John Kingston, Daniel Mash, Della Chambless, and Shigeto Kawahara (Linguist. Dept., Univ. of Massachusetts, Amherst, MA 01003-9274)

English listeners identify a stop ambiguous between /t/ and /k/ more often as ''k'' after /s/ than /sh/ (Mann and Repp, 1981). Judgments shift similarly after a fricative ambiguous between /s/ and /sh/ when its identity is predictable from a transitional probability bias but perhaps not from a lexical bias (Pitt and McQueen, 1998; cf. Samuel and Pitt, 2003). In replicating these experiments, we add a discrimination task to distinguish between the predictions of competing explanations for these findings: listeners respond ''k'' more often after /s/ because they compensate for the fronting of the stop expected from coarticulation with /s/ or because a stop with an F3 onset frequency midway between /t/'s high value and /k/'s low value sounds lower after the /s/'s high-frequency energy concentration. The second explanation predicts listeners will discriminate /s-k/ and /sh-t/ sequences better than /s-t/ and /sh-k/ sequences because sequential contrast exaggerates the spectral differences between /s-k/'s high-low intervals and /sh-t/'s low-high intervals and distinguishes them more perceptually than /s-t/'s high-high intervals and /sh-k/'s low-low intervals. Compensation for coarticulation predicts no difference in discriminability between the two pairs because it does not exaggerate differences between the two intervals. [Work supported by NIH.]

4aSC18. Acoustics and perception of coarticulation at a distance. Karen Jesney, Kathryn Pruitt, and John Kingston (Ling. Dept., Univ. of Massachusetts, Amherst, MA 01003-9274)

CVC syllables were recorded from two speakers of American English, in which the initial and final stops ranged over /b,d,g/ and the vowel ranged over /i,I,e,E,ae,u,U,o,O,a/. F2 locus equations differed systematically as a function of the place of articulation of the other stop. These equation's slope and *y* intercepts were used to synthesize initial /g-b/ and /g-d/ continua in CVC syllables in which the final stop ranged over /b,d,g/ and the vowel over /e,o,a/, and the resulting stimuli were presented to listeners for identification. Listeners responded g significantly more often to both continua when the final stop was /d/ rather than /b/; the number of g responses fell between the /d/ and /b/ extremes for final /g/. This difference between final /d/ vs. /b/ is only observed when the intervening vowel is back /o,a/ and is actually reversed weakly when it is front /e/. Listeners also respond g significantly more often when the final stop is /g/ rather than /b/ and the vowel is /o,a/ but not [e]. Segments do coarticulate at a distance, listeners take this coarticulation into account, and perceptual adjustments depend on the segments through which the coarticulation is expressed. [Supported by NIH.]

4aSC19. Diphones, lexical access, and the verbal transformation effect. James A. Bashford, Jr., Richard M. Warren, and Peter W. Lenz (Dept. of Psych., Univ. of Wisconsin—Milwaukee, P.O. Box 413, Milwaukee, WI 53201-0413, bashford@uwm.edu)

When listeners are presented with a repeating verbal stimulus, adaptation occurs and perception of the stimulus is replaced by perception of a competitor. The present study examines the first of these verbal transformations reported by 180 listeners who were presented with lexical and nonlexical consonantvowel (CV) syllables that varied in frequency-weighted neighborhood density (FWND). These stimuli were produced by pairing the six English stop consonants with a set of three vowels. As expected, the majority of initial illusory forms (78%) were neighbors, differing from the stimulus by a single phoneme, and the proportion of lexical neighbors increased with stimulus FWND. Interestingly, FWND had opposite effects upon the lability of consonants and vowels: There was a strong positive correlation [$r = 0.79$, $F(17) = 26.2$, $p < 0.0001$] between FWND and the number of consonant transformations, and in contrast, there was a strong negative correlation [$r = -0.78$, $F(17) = 24.9$, $p < 0.0001$] between FWND and the number of vowel transformations. The implications of these and other findings with these simple diphones will be discussed in relation to current activation-competition theories of spoken word recognition. [Work supported by NIH.]

4aSC20. Acoustic analysis and perceptual evaluation of nasalized /g/ consonant in continuous Japanese. Hisao Kuwabara (Teikyo Univ. of Sci. & Technol., Uenohara, Yamanshi 409-0193, Japan)

It is well known that the /g/ consonant, a velar voiced plosive, in Japanese continuous speech is often nasalized unless it appears at the word-initial position. The nasalized /g/ consonant takes place in dialects mainly spoken in northern districts including the Tokyo area where the standard Japanese is spoken. However, the number of nasalized /g/ consonants is said to be decreasing year by year according to a survey. This paper deals with acoustic and perceptual analysis of this phenomenon. Test materials used in this experiment are read version of Japanese short sentences by NHK's (Japan Broadcasting Corporation) professional announcers. Each sentence includes at least one /g/ consonant that would likely be pronounced as nasalized. An evaluation test reveals that less than 60% of nasalization has been found to occur for /g/ consonants for which 100% nasalization had been observed decades ago. Acoustic analysis for nasalized and non-nasalized /g/ sounds has been performed mainly through waveform parameters. It has been found that the power ratio between consonant and vowel is the most effective parameter for distinguishing nasals from non-nasals, but it is highly speaker dependent.

4aSC21. Production and perception of place of articulation errors. Adrienne M. Stearns and Stefan A. Frisch (Univ. of South Florida, 4202 E. Fowler Ave., PCD1017, Tampa, FL 33620)

Using ultrasound to examine speech production is gaining popularity because of its portability and noninvasiveness. This study examines ultrasound recordings of speech errors. In experiment 1, ultrasound images of participants' tongues were recorded while they read tongue twisters. Onset stop closures were measured using the angle of the tongue blade and elevation of the tongue dorsum. Measurements of tongue twisters were compared to baseline production measures to examine the ways in which erroneous productions differ from normal productions. It was found that an error could create normal productions of the other category (categorical errors) or abnormal productions that fell outside the normal categories (gradient errors). Consonant productions extracted from experiment 1 were presented auditory-only to naive listeners in experiment 2 for identification of the onset consonant. Overwhelmingly, the participants heard normal productions and gradient error productions as the intended sound. Categorical error productions were judged to be different from the intended sound. The only effect of erroneous production on perception appears to be a slight increase in reaction time, which may suggest that error tokens are abnormal in some way not measured in this study.

4aSC22. Role of linguistic experience on audio-visual perception of English fricatives in quiet and noise backgrounds. Yue Wang, Haisheng Jiang, Chad Danyluck (Dept. of Linguist., Simon Fraser Univ., Burnaby, BC, V5A 1S6, Canada), and Dawn Behne (Norwegian Univ. of Sci. and Technol., Trondheim, Norway)

Previous research shows that for native perceivers, visual information enhances speech perception, especially when auditory distinctiveness decreases. This study examines how linguistic experience affects audio-visual (AV) perception of non-native (L2) speech. Native Canadian English perceivers and Mandarin perceivers with two levels of English exposure (early and late arrival in Canada) were presented with English fricative-initial syllables in a quiet and a caf-noise background in four ways: audio-only (A), visual-only (V), congruent AV, and incongruent AV. Identification results show that for all groups, performance was better in the congruent AV than A or V condition, and better in quiet than in caf-noise background. However, whereas Mandarin early arrivals approximate the native English patterns, the late arrivals showed poorer identification, more reliance on visual information, and greater audio-visual integration with the incongruent AV materials. These findings indicate that although non-natives were more attentive to visual information, they failed to use the linguistically significant L2 visual cues, suggesting language-specific AV processing. Nonetheless, these cues were adopted by the early arrivals who had more L2 exposure. Moreover, similarities across groups indicate possible perceptual universals involved. Together they point to an integrated network in speech processing across modalities and linguistic backgrounds. [Work supported by SSHRC.]

4aSC23. Voicing of /h/ in the Texas Instrument MIT (TIMIT) database. Laura Koenig (Haskins Labs, 300 George St., New Haven, CT 06511 and Long Island Univ.-Brooklyn)

Although English /h/ is traditionally described as voiceless, authors have long recognized that voiced allophones exist, especially in unstressed, intervocalic positions. In past work, we have suggested that fully voiced /h/ may be more common in men than women, but our subject population was limited in number and dialectal diversity. In this study, we use the TIMIT database to obtain measures of /h/ production in men and women speaking multiple dialects of American English. Our analysis focuses on the /h/ initiating the word "had" in a sentence produced by all speakers in the database: "She had your dark suit . . . " Each token of /h/ is coded for whether (a) the /h/ is deleted (i.e., auditorily imperceptible); and, if /h/ is present, whether (b) voicing continues unbroken and (c) there is visible aspiration noise in the speech signal. This analysis will provide baseline data on /h/ realization in a common sentence context. We will also carry out follow-up analyses on selected utterances to gain more insight into the effects of phonetic context, stress, and lexical type (e.g., content versus function word) on the characteristics of English /h/. [Work supported by NIH.]

4aSC24. On distinctions between stops and similar-place weak fricatives. James M. Hillenbrand (Speech Pathol. and Audiol., MS5355, Western Michigan Univ., Kalamazoo, MI 49008) and Robert A. Houde (Ctr. for Commun. Res., Rochester, NY 14623)

There is an extensive body of literature on the acoustic properties of both stops and fricatives. However, little attention has been paid to the acoustic features that might distinguish these two sound categories. This is unsurprising in the case of stops versus strong fricatives since these sounds are rarely confused. Stops and weak fricatives, on the other hand, are frequently confused [G.A. Miller and P.E. Nicely, J. Acoust. Soc. Am. **27**, 338–352 (1955)]. The present study was undertaken in a search for acoustic features that might distinguish the stop/weak fricative pairs /b/–/v/ and /d/–/dh/ (i.e., /d/ vs voiced th). Speech material consisted of CV and VCV syllables spoken by five men and five women, using the vowels /a/, /i/, and /u/. A combination of two features reliably separated

the stops from weak fricatives: (1) intensity during the consonant occlusion interval (typically greater for the fricatives), and (2) the rate of increase in mid- and high-frequency energy (above 1 kHz) associated with consonant release (typically greater for the stops).

4aSC25. Salience of virtual formants as a function of the frequency separation between spectral components.

Robert A. Fox, Ewa Jacewicz, Chiung-Yun Chang, and Jason D. Fox (Speech Acoust. and Percpt. Labs, Dept. of Speech and Hearing Sci., Ohio State, 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210-1002)

The center-of-gravity (COG) hypothesis proposed by Chistovich and others for the perception of static vowels suggests that auditory spectral integration may occur when two or more formants fall within a 3.5 bark bandwidth. While several studies have examined the bandwidth limits of such integration, this study examines the extent to which spectral integration is uniform within this putative 3.5-bark range. We examine the perceptual salience of virtual formants produced by modifying the spectral COG of two closely spaced narrow-bandwidth resonances. Three different vowel series were created: [i-ɪ], [ʌ-ɑ] and [ə-ɚ]. A second set of vowels was then created in which one of the formants (F1 in [i-ɪ], F2 in [ʌ-ɑ] and F3 in [ə-ɚ]) was replaced by a virtual formant whose COG matched that of the formant that had been removed. The frequency separation between the two component resonances was then systematically varied between 1.5 and 3.5 barks and a singleinterval 2AFC vowel identification task was used to obtain estimates of vowel quality for each series step. Results will be discussed in terms of whether the spectral integration effects within the 3.5 bark decline as the frequency separation between the resonance components increases. [Work supported by NIDCD R01DC00679-01A1.]

4aSC26. Frequency effects in phoneme processing.

Danny R. Moates, Noah E. Watkins, Zinny S. Bond, and Verna Stockmal (Inst. for the Empirical Study of Lang., Ohio Univ., Athens, OH 45701, moates@ohio.edu)

Are phonological segments activated during word recognition in proportion to their frequency of use? Previous evidence for this hypothesis [Moates *et al.*, *Laboratory Phonology 7*, edited by Gussenhoven and Warner (Mouton de Gruyter, 2002)] used a word reconstruction task. The present study used an online task, the gating task, in which progressively longer fragments of a word are presented to listeners who must identify the word in as few gates as possible. High- and low-frequency segments were contrasted by presenting them in word pairs that differed only in those two segments, e.g., label versus cable, where /l/ is used more frequently than /k/ [M. K. Vitevich and P. A. Luce, Behav. Res. Methods, Instrum. Comput. **36**, 481–487 (2004)]. We constructed 30 pairs of two-syllable words for which the contrasting segments were at the first syllable onset, 30 more for the second syllable onset, and 30 more for the coda of a syllable. Identification judgments were gathered from 120 participants. *t* tests showed high-frequency segments to be identified at significantly earlier gates than their matched low-frequency segments for first onset and coda, but not second onset. These results offer further evidence for sublexical processes in spoken word recognition.

4aSC27. The clustering coefficient of phonological neighborhoods influences spoken word recognition.

Michael Vitevitch (Dept. of Psych., Univ. of Kansas, 1415 Jayhawk Blvd., Lawrence, KS 66045-7556)

Neighborhood density refers to the number of words, or neighbors, that are phonologically related to a given word. For example, the words BAT, MAT, CUT, and CAN (among others) are considered phonological neighbors of the word CAT. In contrast, the clustering coefficient of the neighborhood refers to the proportion of phonological neighbors that are also neighbors of each other. Among the neighbors of CAT, the words BAT and MAT are neighbors of each other, but the words BAT and CAN are not neighbors of each other. Despite the stimulus words having the same number of neighbors overall, the results of an auditory lexical decision task showed that words with a high clustering coefficient (i.e., most neighbors were also neighbors of each other) were responded to more quickly than words with a low clustering coefficient (i.e., few neighbors were also neighbors of each other). These results suggest that some aspects of phonological similarity (i.e., clustering coefficient) might facilitate lexical activation, whereas other aspects of phonological similarity (i.e., neighborhood density) influence a later, decision stage of processing characterized by competition among activated word-forms. [Work supported by NIH.]

4aSC28. The influence of noise and reverberation on vowel recognition: Response time.

Magdalena Blaszak and Leon Rutkowski (Div. of Rm. Acoust. and Psychoacoustics, Adam Mickiewicz Univ., Umultowska 85, 61-614 Poznan, Poland)

This study examines the perceptual effect of two types of noise and reverberation on vowel recognition. Multitalker babble and traffic noise (European Standard EN 1793–3) were generated simultaneously with Polish vowels /a, e, i, o, u, y/ in two different sound fields and an anechoic chamber. The experiment was performed under various conditions of signal-to-noise ratio (-9, -6, -3, 0, $+3$, no noise). A new procedure for listeners' selection based on the Bourdon's psychometrics test was proposed. The effects of noise and reverberation were quantified in terms of (a) vowel recognition scores for young normal-hearing listeners (YNH) and (b) *ease of listening* based on the time of response and subjective estimation of difficulty. Results of the experiment have shown that (a) the response time can be a good measure of the effect of noise and reverberation on the speech intelligibility is rooms, and (b) in this type of experiment, of great significance is the choice of the subjects based on the psychometric tests.

4aSC29. Quantifying the benefits of sentence repetition on the intelligibility of speech in continuous and fluctuating noises.

Isabelle Mercille, Roxanne Larose, Christian Giguère, and Chantal Laroche (Univ. of Ottawa, 451 Smyth Rd., Ottawa, ON, Canada K1H 8M5)

Good verbal communication is essential to ensure safety in the workplace and social participation during daily activities. In many situations, speech comprehension is difficult due to hearing problems, the presence of noise, or other factors. As a result, listeners must often ask the speaker to repeat what was said in order to understand the complete message. However, there has been little research describing the exact benefits of this commonly used strategy. This study reports original data quantifying the effect of sentence repetition on speech intelligibility as a function of signal-to-noise ratio and noise type. Speech intelligibility data were collected using 18 normal-hearing individuals. The speech material consisted of the sentences from the Hearing In Noise Test (HINT) presented in modulated and unmodulated noises. Results show that repeating a sentence decreases the speech reception threshold (SRT), as expected, but also increases the slope of the intelligibility function. Repetition was also found to be more beneficial in modulated noises (decrease in SRT by 3.2 to 5.4 dB) than in the unmodulated noise (decrease in SRT by 2.0 dB). The findings of this study could be useful in a wider context to develop predictive tools to assess speech comprehension under various conditions.

4aSC30. The effect of the spectral shape changes on voice perception.

Mika Ito, Bruce R. Gerratt, Norma Antonanzas-Barroso, and Jody Kreiman (Div. of Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, jkreiman@ucla.edu)

Researchers have long known that the shape of the vocal source spectrum is an important determinant of vocal quality, but the details regarding the importance of individual spectral features remains unclear. Previous research indicates four spectral features, H1-H2, the spectral slope above 4 kHz, the slope from 1.5–2 kHz, and the slope from 2–4 kHz, account for

virtually all the variability in spectral shapes. The present study provides preliminary evidence about the perceptual importance of these four features. Four series of stimuli were synthesized for each spectral parameter, in which that parameter varied in small steps. Because the perceptual salience of source parameters depends on F0 and on the spectrum of the inharmonic part of the source, series differed in the sex of the speaker (male/female) and in the NSR (noise-free/very noisy). Listeners heard all possible pairs of voices within each series and were asked to determine whether stimuli were same or different. We hypothesize that listeners sensitivity to H1-H2 and the slope of the spectrum from 1.5–2 kHz will be independent of noise, but that sensitivity to changes in the spectral shape above 2 kHz will depend on the amount of noise excitation present in the voice.

4aSC31. The use of auditory and visual information in the perception of stress in speech. James Harnsberger, Daniel Kahan, and Harry Hollien (Inst. for Adv. Study of the Commun. Proc., Univ. of Florida, Gainesville, FL 32611)

Prior work on the acoustic correlates of the perception of psychological stress in speech has suffered from the problem of quantifying and verifying the extent to which a speaker was under stress during articulation. Two experiments were conducted to address this issue. First, stressed and unstressed speech samples were elicited from 78 speakers of American English. Stressed samples were recorded by having subjects read a standard passage while under the threat of the administration of mild electric shock. Both visual and audio recordings were collected. Stress was quantified in terms of four measures: two physiological (pulse rate and galvanic skin response) and two self-report scales. Sentences from the 16 speakers showing the largest differences between the stressed and unstressed conditions were then presented in a paired comparison task to 90 naive listeners, 30 each in three conditions: (1) audio-only presentation of the stimuli, (2) visual-only presentation of the stimuli, and (3) audiovisual presentation of the stimuli. The results indicate that individual listeners are sensitive to stress cues in speech in all three conditions.

4aSC32. Selective attention and perceptual learning of speech. Alexander L. Francis (Dept. of Speech, Lang. & Hearing Sci., Purdue Univ., West Lafayette, IN 47907), Natalya Kaganovich, and Courtney Driscoll (Purdue Univ., West Lafayette, IN 47907)

Phonetic experience can change the perceptual distance between speech sounds, increasing both within-category similarity and between-category distinctiveness. Such warping of perceptual space is frequently characterized in terms of changes in selective attention: Listeners are assumed to attend more strongly to category-differentiating features while ignoring less relevant ones. However, the link between the distribution of selective attention and categorization-related differences in perceptual distance has not been empirically demonstrated. To explore this relationship, listeners were given 6 h of training to categorize sounds according to one of two acoustic features while ignoring the other. The features were voice onset time and onset $f0$, which are normally correlated and can both serve as a cue to consonant voicing. Before and after training, listener's performance on a Garner selective attention task was compared with assessment of the perceptual distance between tokens. Results suggest that training can induce both warping of perceptual space and changes in the distribution of selective attention, but the two phenomena are not necessarily related. Results are consistent with a two-stage model of perceptual learning, involving both preattentive adjustment of acoustic cue weighting and higher-level changes in the distribution of selective attention between acoustic cues. [Work supported by NIH-NIDCD 1R03DC006811.]

4aSC33. Investigation of interaction between speech perception and production using auditory feedback. Masato Akagi, Jianwu Dang, Xugang Lu, and Taichi Uchiyamada (School of Information Sci., JAIST, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan)

This study employed an auditory feedback paradigm with perturbed fed-back speech to investigate interaction between speech perception and production by measuring simultaneous fluctuations of speech production organs using the electromyographic (EMG) signals, articulatory movements, as well as spectral analyses, where the articulatory data were obtained by the electromagnetic articulographic (EMA) system. Chinese vowels pair [i]-[y] and Japanese vowels pairs [e]-[a], [e]-[i] and [e]-[u] were chosen as the experimental objects. When the speaker is maintaining the first vowel, the feedback sound is randomly changed from the first vowel to the second one in each pair by manipulating the first three formants. Spectral analysis showed that a clear compensation was seen in the first and second formants of the vowels. Analyses of EMG and EMA signals also showed muscle reactivation and tongue movements to compensate for the perturbations. Latency of the compensating response is about 150 ms to start and about 290 ms for maximum compensation from the onset of the perturbation. According to the measurements, it seems that in most cases the speaker attempts to compensate for the ''error'' caused by the auditory perturbation by a real-time monitoring, and the auditory feedback takes place simultaneously often during speech production.

4aSC34. Cross-ear suppression of the verbal transformation effect: Tweaking an acoustic-phonetic level. Peter W. Lenz, James A. Bashford, Jr., and Richard M. Warren (Dept. of Psych., Univ. of Wisconsin—Milwaukee, P.O. Box 413, Milwaukee, WI 53201-0413, plenz@uwm.edu)

A recorded word repeating over and over undergoes a succession of illusory changes. When two images of the same repeating word are presented dichotically, with a half-cycle delay preventing fusion, the two images of the word each undergo independent illusory transformations at a rate equivalent to that of a single image [Lenz *et al.*, J. Acoust. Soc. Am. **107**, 2857 (2000)]. However, with one phoneme difference (e.g., ''dark'' versus ''dart''), transition rate is dramatically suppressed [Bashford *et al.*, J. Acoust. Soc. Am. **110**, 2658 (2001)]. Rates decrease with extent of feature mismatch at a single phoneme position (roughly 30% reduction with one feature mismatch and 45% with three). Rates also decrease with the number of mismatched phonemes (about 80% rate reduction with three out of four), suggesting a strong acoustic-phonetic basis for verbal transformation suppression. In contrast, semantic relations had no effect (e.g., transformations for ''light'' were suppressed equally by contralateral night and ''might''). Dichotic competition appears to allow us to access and selectively influence a prelexical stage of linguistic analysis. [Work supported by NIH.]

4aSC35. Perceptually balanced filter response for binaural dichotic presentation to reduce the effect of spectral masking. Pandurangarao N. Kulkarni, Prem C. Pandey (Elec. Eng. Dept, Indian Inst. of Technol. Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in), and Dakshayani S. Jangamashetti (Basaveshwar Eng. College Bagalkot, Bagalkot Karnataka 587102, India)

Earlier investigations show that the scheme of binaural dichotic presentation with spectral splitting of speech signal helps in reducing the effect of spectral masking for persons with moderate bilateral sensorineural hearing impairment. Speech perception improved by employing filters with interband crossover gain adjusted between 4 and 6 dB below the pass band gain. The relationship between scaling factors for a tone presented to the two ears, so that perceived loudness is that of a monaural presentation, is investigated for design of comb filters with improved perceptually balanced response. Results from the listening tests show that, for perceptual balance, the sum of the two scaling factors should be constant, indicating that the magnitude response of the comb filters should be complementary on a linear scale.

4aSC36. The organization of bilingual perceptual consonant space: English/Spanish bilingual perception of Malayalam nasal consonants. Jenna Silver and James Harnsberger (Inst. for Adv. Study of the Commun. Proc., Univ. of Florida, Gainesville, FL 32611)

This study examines the capacity of English/Spanish bilinguals to discriminate between consonants that exist in only one of their respective phonetic inventories. Two non-native nasal consonant contrasts were tested: dental versus alveolar and the palatal versus velar, both found in Malayalam. The dental and palatal nasals appear in Spanish, while the alveolar and velar nasals occur in English. Poorer performance in discrimination was interpreted as indicative of a common nasal category subsuming the Spanish dental and English alveolar nasals; better performance was taken as evidence of the maintenance of separate categories from both languages. Two other tests were administered to aid in the interpretation of the discrimination test scores: forced-choice identification and perceptual similarity ratings. The findings of this research will be used to characterize the perceptual consonant space in terms of continuum between two possible bilingual systems: one that collapses similar categories across languages or one that maintains two distinct phonological systems that can be accessed simultaneously. It is believed that bilinguals will be able to discriminate between these contrasts more consistently than their monolingual peers; however, there is no prediction about performance relative to the monolingual group from Malayalam.

4aSC37. Agreement and reliability using reference-matching paradigm in perceptual voice quality rating in Chinese and English. Mandy Ho and Edwin Yiu (Voice Res. Lab., Div. of Speech & Hearing Sci., Univ. of Hong Kong, 5/F Prince Philip Dental Hospital, Hong Kong)

Perceptual voice judgment is commonly used in clinical voice quality evaluation. The use of a referencematching paradigm in perceptual ratings has been shown to improve both agreement and reliability (Yiu *et al.*, in press). This study set out to investigate the agreement and reliability in rating Chinese and English dysphonic stimuli using the reference-matching paradigm. Experiment 1 aimed to synthesize Chinese and English dysphonic stimuli with different breathy and rough severity levels using the HLSyn Speech Synthesizer. Seven representative anchors (references) for each of the rough and breathy series in Chinese and English were chosen by three judges to be used in experiment 2. Acoustic analysis of the anchor series indicated they were of increasing severity. Experiment 2 recruited ten native Chinese and ten native English subjects to rate the quality of Chinese and English dysphonic voice samples using the synthesized anchor as references. Results showed that listeners achieved nearly 90% agreement in rating the Chinese stimuli and 80% agreement in rating the English stimuli regardless of their language background. The study showed that the reference-matching paradigm was a reliable method in rating dysphonic stimuli across listeners with different language backgrounds.

4aSC38. Learning to perceive non-native speech sounds: The role of test stimulus variability. McNeel Jantzen and Betty Tuller (Ctr. for Complex Systems and Brain Sci., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431)

Natural speech stimuli used in studies of phonological learning usually include several in talkers and phonetic environments because variability aids learning [e.g., Lively, Logan, and Pisoni, J. Acoust. Soc. Am. (1993)]. The present study investigated whether nonphonetic variability in the synthetic test set has a similar effect. First, a perceptual mapping procedure was performed using a synthetic continuum that ranged from the Malayalam voiced, unaspirated, dental stop consonant to the American English alveolar [d], with three $F0$ contours (low, mid, and high). Subjects identified the stimuli (2AFC) and judged their goodness as exemplars of each category. Subjects then received 15 sessions (one/day) of 2AFC training with feedback using natural stimuli produced by native Malayalam speakers, and performed difference ratings on a subset of pairs from the syn-

thetic stimuli. The perceptual mapping procedure was repeated at 1 and 14 days post-training and results compared with a parallel experiment that included only the midlevel $F0$ contour in the synthetic test set. [Work supported by NSF.]

4aSC39. Influence of the prosody of spoken language on recognition and memory for vocal quality. Sumi Shigeno (Aoyama Gakuin Univ., 4-4-25 Shibuya, Shibuya-ku, Tokyo, 150-8366 Japan)

This study examined whether recognition and memory for vocal quality of a speaker who speaks either a native language or non-native language should be influenced by the prosody of the language that the speaker utters. Voices of 12 speakers were recorded. They were six Japanese people and six Americans and Britons. All speakers uttered short sentences in their respective native languages (i.e., Japanese for Japanese speakers and English for Americans and Britons) and in a non-native language (i.e., English for Japanese speakers and Japanese for Americans and Britons). Ten Japanese participants rated the vocal quality of speakers in the first session. After 1 week the same experiment was again conducted in the second session. Results showed that the performance of identification of speakers as Japanese or as non-Japanese was comparatively accurate even though the ratings on the speakers' voices were varied as the language spoken by the speakers. Ratings of the voices were compared further between two sessions and little difference was found, irrespective of a 1-week blank. Results suggest that the memory for vocal quality is robust, but that the recognition of vocal quality is dependent on the prosody of the language spoken by speakers.

4aSC40. Brain activity during auditory processing affected by expectation of speech versus nonspeech. Yukiko Nota (ATR CIS BAIC, 2-2-2 Hikaridai, Keihanna Sci. City, Kyoto 619-0288, Japan, ynota@atr.jp)

fMRI was used to clarify whether there is any differential brain activity invoked by expectation for speech versus nonspeech sounds. Auditory stimuli were created by acoustically morphing between either sustained vowels or tones, respectively, and a buzz sound. The two sets of interpolation were performed in nine nonlinear steps; the stimuli retained for perceptual experiments were only the three most vowel-like, the three most tone-like, and the three most buzz-like tokens morphed from the vowels. In the "speech expectation" session, subjects were instructed to discriminate the vowel-like and buzz-like stimuli; in the "nonspeech expectation" session, subjects were instructed to discriminate the tone-like and buzz-like stimuli without knowing that the buzz stimuli had been morphed from the vowels. Thus the buzz-like stimuli in both experiments were the same, but the subjects' expectations were different because they were told to expect either speech (vowel-like) or nonspeech (tone-like) stimuli. Comparison of brain activation during processing of the buzz-like stimuli under these two conditions revealed that BA40 and thalamus were more activated in speech expectation, while right BA20 was more activated in nonspeech expectation. These results suggest that subjects' speech/nonspeech expectation for sound stimuli influences brain activity for actual auditory processing. [Work supported by MEXT.]

4aSC41. Representations involved in short-term versus long-term word learning by preschool children with and without phonological disorders. Holly Storkel, Jill Hoover, and Junko Maekawa (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, 1000 Sunnyside Ave., 3001 Dole Ctr., Lawrence, KS 66045-7555, hstorkel@ku.edu)

This study explores whether sublexical (i.e., individual sound) and/or lexical (i.e., whole-word) representations contribute to word learning and whether these contributions change across short-term versus long-term learning. Sublexical representations were indexed by phonotactic probability, the likelihood of occurrence of a sound sequence, whereas lexical representations were indexed by neighborhood density, the number of

similar sounding words. Thirty-four preschool children participated in a short-term word learning task that exposed them to nonwords varying in phonotactic probability and neighborhood density and tested learning of these nonwords. Long-term learning was assessed through comprehension and production of real words varying in phonotactic probability and neighborhood density. Results showed that phonotactic probability and neighborhood density equally influenced short-term word learning. In contrast, long-term word learning was affected primarily by neighborhood density. Thus, both sublexical and lexical representations appeared to play a role in short-term learning, but only lexical representations played a primary role in long-term retention. This pattern held for both children with normal phonological development and children with phonological delays. However, the direction of the effect of neighborhood density for short-term word learning varied by group status, suggesting differences in the use of lexical representations during short-term learning. [Work supported by NIH.]

4aSC42. Changes in formant frequencies associated with postural change in adult male speakers over 50 years old. Michiko Hashi, Tomoki Nanto, and Natsuki Ohta (Prefectural Univ. of Hiroshima, 1-1 Gakuen-cho, Mihara, Hiroshima, Japan)

It is possible that changes of direction of gravity relative to the vocal tract associated with changes in posture influence acoustic characteristics of speech including vowel formant frequencies. Studies examining such effects had produced mixed results and demonstrated the possibility of substantive interspeaker variability in the effect of postural changes on vowel formant frequencies. Recent work by Takakura *et al.* ["Changes in formant frequencies associated with postural change," paper presented at the Fall meeting of Acoustical Society of Japan (2006)], using young adult male speakers, revealed a small number of speakers demonstrating changes in vowel formant frequencies and suggested effect of age. The present study attempts to examine changes of vowel formant frequencies in upright and supine position among older male speakers. Attempts will be made to eliminate the effect of differences in neck position between the postures through the use of a power-bead-based neck stabilizer. The results will be compared with data from young normal speakers in the previous study and inferences will be made relative to speech production models.

4aSC43. The effect of viewing angle on the visual contribution to speech intelligibility in noise. Eugene Brandewie, Douglas Brungart, Nandini Iyer, and Brian Simpson (Air Force Res. Lab., Wright–Patterson AFB, Ohio 45433-7901)

Visual cues are known to assist speech comprehension in noisy environments, but relatively little is known about the impact that viewing angle has on the visual contribution to speech intelligibility. In this experiment, four digital cameras were used to make simultaneous recordings of test phrases from the Modified Rhyme Test at four different viewing angles: 0, 45, 90, and 135 deg. These test phrases were used to measure the effect of viewing angle on the intelligibility of noise-masked speech stimuli that were presented with and without visual cues at seven different signal-to-noise ratios (SNRs). When the face was viewed from the front, the visual cues provided an intelligibility improvement that was equivalent to a 6–10-dB increase in SNR. This visual benefit remained roughly constant for viewing angles up to 90 deg, but it dropped off rapidly (to less than 2 dB) when the viewing angle increased to 135 deg. The results suggest that the visual contribution to speech perception is severely impaired when the observer does not have access to an unobstructed view of the talker's mouth.

4aSC44. Towards estimation of Japanese intelligibility scores using objective voice quality assessment measures. Rui Kaga, Kazuhiro Kondo, Kiyoshi Nakagawa, and Masaya Fuzimori (Yamagata Univ., Jounan 4-3-16, Yonezawa, 992-8510, Yamagata, Japan)

We investigated the use of objective quality measures to estimate the intelligibility of Japanese speech. We initially focused on PESQ (perceptual evaluation of speech quality), which is the state-of-the-art objective assessment method, and can estimate the mean opinion scores (MOS) at an extremely high accuracy. Since we can assume that speech quality is correlated with intelligibility, it should be possible to estimate the intelligibility from the estimated opinion scores or some of its derivatives. We tried to estimate the intelligibility of the Japanese Rhyme Test (DRT). The DRT uses minimal word pairs whose initial phone differs by one phonetic feature. We estimated the MOS of the DRT word samples mixed with noise and tried to correlate this with the measured intelligibility. However, the estimated MOS showed no difference between phonetic features. However, the difference in the estimated MOS between the word pairs seems to differ by a phonetic feature for SNR above 10 dB, which suggests that the estimation of intelligibility by a phonetic feature may be possible. We also plan to selectively use the internal data used to calculate the MOS estimate for better intelligibility estimation.

Session 4aSP

Signal Processing in Acoustics, Underwater Acoustics, and Acoustical Oceanography: Adaptive Signal Processing

Juan I. Arvelo, Jr., Cochair

Johns Hopkins Univ., Applied Physics Lab., National Security Technology Dept., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099

Kensaku Fujii, Cochair

Univ. of Hyogo, School of Engineering and Graduate School of Engineering, 2-67 Shosha, Himeji, Hyogo 671-2201, Japan

Chair's Introduction—8:00

Invited Papers

8:05

4aSP1. Adaptive beamforming for multipath environments. Henry Cox (Lockheed Martin, IS&S, AC3DI, 4350 North Fairfax Dr., Ste. 470, Arlington, VA 22203)

Coherent multipaths present a significant challenge to most adaptive beamformers because they violate the common assumption of a rank-one plane wave or geometrically focused signal. When the multipath arrivals that characterize shallow water propagation are resolvable by the array's aperture, the mismatch between the assumed and the true signal spatial structure causes signal suppression. If the amplitude and phase relationships among the various multipaths were known, they could, in principle, be included in a matched field beamforming approach. This is usually impractical due to inadequate knowledge of propagation parameters, especially bottom characteristics, and source/receiver motion. A generalization of the standard MVDR approach, called multirank MVDR, assumes only that the signal lies in a subspace of multiple rank rather than the usual rank-one assumption. An example of a subspace is a small fan of beams that cover the potential multipath directions. The signal may be rank one corresponding to fully coherent multipath or higher rank corresponding to incoherent or partially coherent multipath. The multirank approach is applied to the shallow water multipath problem and compared with the related technique of employing multiple linear constraints. Realistic simulations of alternative beamforming approaches for a large horizontal array are presented. [Work supported by ONR.]

8:25

4aSP2. Robust adaptive algorithm based on nonlinear error cost function for acoustic echo cancellation. Suehiro Shimauchi, Yoichi Haneda, and Akitoshi Kataoka (NTT Cyber Space Labs., NTT Corp., 3-9-11, Midori-cho, Musashino-shi, Tokyo, 180-8585, Japan, shimauchi.suehiro@lab.ntt.co.jp)

Motivated by recent progress in the blind source separation (BSS) technique, a robust echo cancellation algorithm is investigated, which would inherently identify the echo path even during double-talk by separating the acoustic echo from the local speech. An adaptive filter has been introduced into acoustic echo cancellers to identify the acoustic echo path impulse response and generate the echo replica. However, most adaptive algorithms suffer from instability during double-talk. Although step-size control cooperating with a double-talk detector (DTD) is a promising approach to stop the adaptation temporarily during double-talk, it cannot handle the echo path change during double-talk. To overcome this problem, novel robust algorithms are derived by applying nonlinearity to the cost function of a conventional echo cancellation algorithm such as the normalized least mean squares algorithm (NLMS) or the affine projection algorithm (APA). Using the simulation results, there is a discussion about how the robustness of the derived algorithms depends on the choice of the nonlinear function and the original algorithm.

8:45

4aSP3. Realistic modeling of adaptive beamformer performance in nonstationary noise. Bruce K. Newhall (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723)

Most adaptive beamformers (ABFs) operate under the assumption that the noise field is quasistationary. They estimate the present noise field by averaging, assuming stationarity over the estimation time. The adaptive beamformer responds to slow changes in the noise field across multiple estimation intervals. Unfortunately, in many low-frequency underwater sound applications, the shipping noise may change rapidly, due to nearby ship motion. This motion can be significant during the estimation interval and degrade ABF performance. A realistic model has been developed, including two effects of source motion on horizontal towed arrays. Bearing rate produces a differential Doppler shift across the array. Range rate produces an amplitude modulation as the multipath interference pattern shifts along the array. The ABF model begins with a realization of ship locations and motion based on historical shipping density. Each ship generates realistic random noise composed of tonals in a broadband background. The noise is propagated from each ship to each hydrophone by a normal mode model. For each time sample the position of each ship is updated and propagation recalculated. The ability of a variety of ABF algorithms to reduce shipping noise clutter is simulated and compared.

4aSP4. Multichannel active noise control system without secondary path models using the simultaneous perturbation algorithm.

Yoshinobu Kajikawa and Yasuo Nomura (Faculty of Enginnering, Kansai Univ., 3-3-35, Yamate-cho, Suita-shi, Osaka 564-8680, Japan, kaji@joho.densi.kansai-u.ac.jp)

This paper presents a novel multichannel active noise control (ANC) system without secondary path models. This ANC system uses a simultaneous perturbation algorithm as the updating algorithm and has an advantage that secondary path models (estimation of secondary paths) are not required, unlike the MEFX (multiple error filtered-X)-based ANC. This system can consequently control noise stably because there are not modeling errors that cause system instability. The computational complexity is also very small. Experimental results demonstrate that the proposed multi-channel ANC system can operate stably under the environment where the error microphones always move.

Contributed Papers

4aSP5. Two-microphone system using linear prediction and noise reconstruction.

Hirofumi Nakano, Kensaku Fujii (Dept. of Comput. Eng., Univ. of Hyogo, 2167 Shosya, Himeji 671-2201, Japan, er06j025@steng.u-hyogo.ac.jp), Tomohiro Amitani, Satoshi Miyata (TOA Corp., Takarazuka Hyogo 665-0043, Japan), and Yoshio Itoh (Tottori Univ., Japan)

This study proposes a new adaptive microphone system that is characterized by a linear prediction circuit inserted previous to a noise reconstruction filter corresponding to an adaptive delay used in conventional systems. This insertion provides various advantages to the adaptive microphone system. One is that steering a null for a noise source becomes possible irrespective of the incidence of speech signal. Another is that the new microphone system works as an omnidirectional microphone for the speech signal. Consequently, setting the microphones at arbitrary intervals is possible. For example, building one microphone in a handset and another into a telephone baseset becomes possible, which provides higher noise reduction effect. In practical use, microphone systems must function in reflective surroundings. In this study, such performance of the proposed system is verified first using computer simulations and then using an experimental system put in an ordinary room. This study also presents experimental results verifying that the proposed system can successfully reduce noise incident from the same direction as a speech signal, as well as crowd noise recorded in an airport.

4aSP6. Adaptive beamformer trade-off study of an expendable array for biologic vocalizations.

Juan Arvelo (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

Adaptive beamformers exploit the ambient noise anisotropy to increase the array gain against this background, enhance target detection, and increase the resolution of the beam response. A prototype expendable array was developed and tested for high-frequency passive detection and localization of marine mammals. This latest array consists of vertical poly-(vinylidene fluoride) (PVDF) wire elements arranged in a horizontal 6 $\times$6 grid with the corner elements removed for a total of 32 lines. The length of the wires forms a vertical beam response that allows exploitation of the ambient noise directionality in elevation while the horizontal aperture provides full coverage in azimuth. The performance and computational demand of selected adaptive and conventional beamformers are compared in a trade-off study to determine their possible use in this expendable system. This trade-off study accounts for the demand of computational resources in addition to the predicted system performance as adjuncts to ocean observatories. [This effort is partly supported by JHU/APL and the Office of Naval Research (ONR).]

4aSP7. Adaptive matched field processing enhancements to forward sector beamforming.

Jeffrey A. Dunne (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723)

A study was undertaken to examine the potential benefit of adaptive matched field processing (AMFP) to the forward sector capability of single-line, twin-line, and volumetric arrays. Comparisons are made with conventional MFP (CMFP) and adapative and conventional plane-wave beamforming (APWB and CPWB) in order to assess the degree of own-ship noise reduction obtainable and any corresponding improvement to the signal-to-noise ratio (SNR). A minimum variance distortionless response beamformer using dominant mode rejection was implemented, applied to both uniform and distorted array shapes. Significant improvement over CMFP and CPWB in tracking and SNR was seen for modeled data in both cases, with the distorted array showing, not surprisingly, better left-right rejection capability. [Work was undertaken with support from the Defense Advanced Research Projects Agency (DARPA) Advanced Technology Office (ATO).]

4aSP8. Vector sensor array sensitivity and mismatch: Generalization of the Gilbert-Morgan formula.

Andrew J. Poulsen and Arthur B. Baggeroer (MIT, Cambridge, MA 02139, poulsen@mit.edu)

The practical implementation of any sensing platform is susceptible to imperfections in system components. This mismatch or difference between the assumed and actual sensor configuration can significantly impact system performance. This paper addresses the sensitivity of an acoustic vector sensor array to system mismatch by generalizing the approach used by Gilbert and Morgan for an array with scalar, omnidirectional elements [E.N. Gilbert and S.P. Morgan, Bell Syst. Tech. J. **34**, (1955)]. As such, the sensor orientation is not an issue because it does not affect performance for an array of omnidirectional sensors. Since vector sensors measure both the scalar acoustic pressure and acoustic particle velocity or acceleration, the sensor orientation must also be measured to place the vector measurement in a global reference frame. Here, theoretical expressions for the mean and variance of the vector sensor array spatial response are derived using a Gaussian perturbation model. Such analysis leads to insight into theoretical limits of both conventional and adaptive processing in the presence of system imperfections. Comparisons of theoretical results and simulations are excellent. One noteworthy result is the variance is now a function of the steering angle. [Work supported by the PLUSNeT program of the Office of Naval Research.]

4a FRI. AM

10:40

4aSP9. Adaptive filtering using harmonic structure of voiced speech for reducing nonstationary known noise. Kenko Ota, Masuzo Yanagida (Doshisha Univ., 1-3, Tarata-Miyakodani, Kyotanabe, Kyoto, 610-0321, Japan, etf1704@mail4.doshisha.ac.jp), and Tatsuya Yamazaki (NICT, 619-0289, Kyoto, Japan, yamazaki@nict.go.jp)

Proposed is an effective method for reducing nonstationary known noise. The objective of this research is to develop a scheme of preprocessing for speech recognition that keeps the same speech recognition rate even in a worse acoustic environment and to realize a TV control system using speech recognition. The basic idea of the proposed method is to estimate the frequency spectrum of noise including sounds from the TV itself and to remove noise components from the frequency spectrum of the received signal. A transfer function from the TV set to the microphone is calculated in an iterative manner estimating the noise signal at the microphone. Traditional ANC techniques do not directly use spectral features such as harmonic structure or fundamental frequency, for example. As the proposed method uses harmonic structure of vocalic segments in command speech, it is expected to have an advantage compared with the traditional adaptive methods. Results of evaluation using speech recognition show that the proposed method significantly improves speech recognition rate compared with conventional spectral subtraction. [Work supported by Knowledge Cluster Project, MEXT, and by Academic Frontier Project Doshisha University.]

10:55

4aSP10. Robust matched-field processor in the presence of geoacoustic inversion uncertainty. Chen-Fen Huang, Peter Gerstoft, and William S. Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92037-0238, chenfen@ucsd.edu)

This presentation examines the performance of a matched-field processor incorporating geoacoustic inversion uncertainty. Uncertainty of geoacoustic parameters is described via a joint posterior probability distribution (PPD) of the estimated environmental parameters, which is found by formulating and solving the geoacoustic inversion problem in a Bayesian framework. The geoacoustic inversion uncertainty is mapped into uncertainty in the acoustic pressure field. The resulting acoustic field uncertainty is incorporated in the matched-field processor using the minimum variance beamformer with environmental perturbation constraints (MV-EPC). The constraints are estimated using the ensemble of acoustic pressure fields derived from the PPD of the estimated environmental parameters. Using a data set from the ASIAEX 2001 East China Sea experiment, tracking performance of the MV-EPC beamformer is compared with the Bartlett beamformer using the best-fit model.

11:10

4aSP11. A study on combining acoustic echo cancelers with impulse response shortening. Stefan Goetze, Karl-Dirk Kammeyer (Dept. of Commun. Univ. of Bremen, Eng., D-28334 Bremen, Germany), Markus Kallinger, and Alfred Mertins (Carl von Ossietzky-Univ., Oldenburg, D-26111 Oldenburg, Germany)

In hands-free video conferencing systems acoustic echo cancelers (AECs) have to face the problem of very high-order impulse responses (IRs), which have to be compensated. Time-domain algorithms for adaptation often suffer from slow convergence (as the NLMS algorithm, e.g.) or high computational complexity (e.g., the RLS). On the other hand frequency-domain algorithms introduce undesired delays [S. Haykin, *Filter Theory*, 2002]. For high-quality hands-free systems IR shortening concepts and IR shaping concepts developed for listening room compensation (LRC) [M.Kallinger and A. Mertins, in Proc. Asilomar, 2005] can be applied to increase speech intelligibility for the near-end speaker. The aim of this study is the synergetic combination of LRC concepts with acoustic echo cancellation. For this scenario two different forms of concatenating the subsystems are possible. Either the AEC filter follows the LRC or vice versa. In the first case the equalization filter reduces the length of the effective IR seen by the AEC filter. Thus, shorter AEC filters can be used which results in faster convergence. However, an estimation algorithm for the room IR is necessary for the LRC subsystem. In the second case the AEC delivers an estimate of the room IR which can be used as an input for the LRC filter. Experimental results confirm the superiority of the new combined approach.

11:25

4aSP12. Krylov and predictive sequential least-squares methods for dimensionality reduction in adaptive signal processing and system identification. James Preisig and Weichang Li (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Rapid time variation of the environment, a large number of parameters which need to be adjusted, and the presence of a reduced subset of the parameters that are relevant at any point in time create significant challenges for adaptive signal-processing algorithms in underwater acoustic applications. In applications such as underwater acoustic communications, the environment is represented by the ''taps'' of the time-varying impulse response. The instability of estimation algorithms or inability to track rapid channel fluctuations are among the problems that are encountered. An approach to addressing these challenges is to dynamically select a subspace in which the adjustment of taps takes place. Here, two algorithms for doing this are presented. The first is based upon using subspace basis vectors, which form a Krylov subspace with respect to the channel input correlation matrix and the channel input/output correlation vector. This method does not use a prediction residual error to select the basis vectors. A second algorithm is a new variant of the matching pursuit algorithm. In this case, ''important'' taps of the channel impulse response are selected to minimize a forward prediction residual error. The properties and performance of these two algorithms are presented and compared using simulation and field data.

11:40

4aSP13. Expectation maximization joint channel impulse response and dynamic parameter estimation and its impact on adaptive equalization. Weichang Li and James C. Preisig (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Joint estimation of channel impulse response and its dynamic parameters using the expectation maximization (EM) algorithm and its MAP variant is derived for broadband shallow-water acoustic communication channels. Based on state-space channel modeling, the EM algorithms estimate the channel dynamic parameters from the sequence of channel impulse response estimates. The estimated parameters are then used in the Kalman smoother, which estimates the channel impulse response. The stability of the algorithm is shown to be related to an extended persistent excitation (EPE) condition, which requires that both the symbol sequence and the channel estimates be persistently exciting. Modified algorithms are proposed for broadband multipath channels to avoid the issue of insufficient excitation. Efficient suboptimal algorithms are also derived from the EM algorithms that alternatively estimate the parameter and the channel impulse response while allowing slow parameter variations. The performance of these channel estimation algorithms as well as their impact on the subsequent equalizer are demonstrated through experimental data analysis. [Work supported by ONR Ocean Acoustics.]

Session 4aUWa

Underwater Acoustics: Sonar Performance

Lisa M. Zurk, Cochair

Portland State Univ., Electrical and Computer Engineering Dept., 1900 S. W. Fourth Ave., Portland, OR 97207

Hiroshi Ochi, Cochair

Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-1-5 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

Chair's Introduction—7:55

Contributed Papers

8:00

4aUWa1. Automatic detection performance comparisons of three different fluctuation-based signal processors. Ronald A. Wagstaff (Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., Univ., MS 38677)

The three most common fluctuation-based signal processor (FBP) algorithms achieve gain by exploiting either the reciprocal of the spectral amplitude, the log-differential amplitude, or aligned-phase angles. Two important features of these processors, for the underwater acoustics community, is their ability to detect and automatically identify signals which originated from submerged sources, and to provide unusually large signal-to-noise ratio gains. Similar benefits are of interest to the atmosphere acoustic community. An example is the automatic detection and identification/classification of hostile airborne and ground vehicles by unattended ground sensors (UGS). The three different generic types of FBP algorithms will be defined. The manner in which each exploits fluctuations to achieve gain will be explained. Corresponding performances will be compared using both underwater and atmosphere acoustic data. The ocean acoustic results came from towed array beamformed spectral data and will include spectral plots and grams. Corresponding single sensor spectral results will also be presented for atmosphere acoustic vehicle data. [Work supported by ARDEC and SMDC.]

8:15

4aUWa2. Evolution of modern fluctuation-based processing. Ronald Wagstaff (Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., Univ., MS 38677)

Modern fluctuation-based processors (FBPs) are relatively new on the signal processing scene. They started in the mid-1980s with the realization that averaging the reciprocal acoustic powers and inverting the result back, i.e., the harmonic mean, could yield 6- to 8-dB signal-to-noise ratio gain over the corresponding average power. Because of its significant noise attenuation, this processor was designated WISPR, i.e., reduces noise to a whisper. WISPR had a unique, potentially more valuable, capability. Based on the decibel difference, or ratio, between the average power and WISPR, it could be determined whether the received signals were from ships, or from sources of sound that were submerged. After much time and experience with WISPR at sea, acquiring and processing towed array ocean acoustic data, and continuing data processing in the laboratory, the phenomena that were responsible for WISPRs performance, acoustic fluctuations generated near the sea surface, became better understood and WISPRs credibility increased. This led to the development of many other FBPs with similar capabilities, but with significantly enhanced performances. A brief account of post-WISPR development will be presented, including a description of the exploitable parameters, how they are used, and the range of gains that they achieve.

8:30

4aUWa3. A comprehensive unbiased third party evaluation of a signal processor for detecting submerged sources among clutter signals and noise. Ronald Wagstaff (Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., Univ., MS 38677)

The Wagstaff's integration silencing processor, WISPR', was developed to detect and identify signals in the ocean from sources that are submerged well below the sea surface. WISPR is the type of signal processor that exploits the reciprocal of the spectral power amplitude, rather than the amplitude as the average power processor does. Processing the reciprocal of the power represented a significant departure in the prevailing signal processing philosophy that governed most conventional signal processing algorithms that were in use when WISPR first appeared on the scene several years ago. Furthermore, WISPR's claimed submerged-source detection capability made it an attractive candidate for some high interest signal processing applications. Accordingly, one influential national organization considered its potential use in their mission and decided to commission a credible third party laboratory to conduct an unbiased evaluation of the WISPR processor. The emphasis was to be on its performance for automatic unalerted detection of signals from submerged sources. The techniques and evaluation methods used to test the WISPR processor will be described. The results of the evaluation will be presented, and the influence of those results on the development of other, more advanced, fluctuation-based processors will be discussed.

8:45

4aUWa4. The estimated ocean detector: Predicted performance for continuous time signals in a random/uncertain ocean. Jeffrey A. Ballard, R. Lee Culver, Leon H. Sibul (Appl. Res. Lab. and Grad. Program in Acoust., Penn State Univ. P.O. Box 30, State College, PA 16804), Colin W. Jemmott, and H. John Camin (Penn State Univ., State College, PA 16804)

This paper addresses implementation of the maximum likelihood (ML) detector for passive SONAR detection of continuous time stochastic signals that have propagated through a random or uncertain ocean. We have shown previously that Monte Carlo simulation and the maximum entropy method can make use of knowledge of environmental variability to construct signal and noise parameter probability density functions (pdf's) belonging to the exponential class. For these cases, the ML detector has an estimator-correlator and noise-canceller implementation. The estimator-correlator detector computes the conditional mean estimate of the signal conditioned on the received data and correlates it with a function of the received data, hence the name estimated ocean detector (EOD). Here we derive the detector structure for continuous time stochastic signals and Gaussian noise and present receiver operating characteristic (ROC) curves for the detector as a function of the signal-to-noise ratio. [Work supported by ONR Undersea Signal Processing Code 321US.]

4a FRI. AM

3259   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3259**

4aUWa5. Echo detection enhancement using multiple guide sources in shallow water. David C. Calvo, Charles F. Gaumond, David M. Fromm, and Richard Menis (Naval Res. Lab., Acoust. Div. Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375)

The use of a guide source has been proposed as a way of compensating for multipath by forming a spatial-temporal cross correlation of the received target and guide source signals across a vertical array in the frequency domain [Siderius *et al.*, J. Acoust. Soc. Am. **102**, 3439–3449]. This processing has the effect of creating a virtual receiver at the guide source position. In general, the performance of a virtual receiver degrades if the spatial integration is not carried out over the span of the array with significant signal. In our study, we have pursued an alternative approach of using guide sources which does not require this integration in general. The guide source signal is simply used as a matched filter. Although this does not correspond to a virtual receiver, it is useful as a means of improving active or passive detection of signals in noise. In general, the signal gain using this alternative technique is dependent on the guide source position. To compensate for this, we construct a separable-kernel-receiver filter bank using multiple randomly positioned guide source signals. Improvement of ROC curves in both passive and active scenarios is obtained using experimental and simulated data. [Work sponsored by ONR.]

4aUWa6. Incorporating environmental variability into received signal statistics. H. John Camin, R. Lee Culver, Leon H. Sibul (Appl. Res. Lab. and Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804), Jeffrey A. Ballard, and Colin W. Jemmott (Penn State Univ., State College, PA 16804)

We have developed a Monte Carlo-based method for estimating the variability of acoustic signal parameters caused by uncertain ocean environments. The method begins with a physics-based model for the environmental properties and uses the maximum entropy (MaxEnt) method to construct probability density functions (pdf's) describing the measured deviations from the model mean. Random realizations of environmental variability, with proper depth correlation, are constructed from the pdf's and added to the mean model parameters. A parabolic equation code (RAM) is used to propagate acoustic energy through each realization of the environment. Fourier synthesis is used to recreate the arrival structure. The method is demonstrated using measurements from the Strait of Gibraltar, which is a particularly complicated region dominated by strong tidal fluctuations and internal waves. During 1996, an international group carried out the Strait of Gibraltar Acoustic Monitoring Experiment (SGAME), in which detailed environmental and 250-Hz acoustic data were collected.

Here, pdf's of the received signal level are compared with results of the Monte Carlo method to demonstrate performance. [Gibraltar data and SVP model provided by Chris Tiemann (ARL:UT) and Peter Worcester (SIO). Work supported by ONR Undersea Signal Processing.]

4aUWa7. Motion compensation of multiple sources. Joung-Soo Park (Agency for Defense Development, P.O. Box18, Jinhae, Kyung-Nam, 645-600, Korea), Jae-Soo Kim (Korea Maritime Univ., Young-Do, Busan, Korea), and Young-Gyu Kim (Agency for Defense Development, Jinhae, Kyung-Nam, 645-600, Korea)

Matched field processing has a advantage of detection of multiple targets. But, if a strong interferer is moving fast near a quiet target, detection of the target is difficult due to the motion effect of the interferer. The motion of the interferer introduces energy spreading and results in poorer detection. A waveguide-invariant-based motion compensation algorithm was proposed to mitigate the motion effect of a dominant signal component, which is estimated by eigenvalue method. The eigenvalue method is good for a strong interferer, but not good for multiple targets. In this presentation, we will propose a steered beam processing method to mitigate the motion effect of multiple targets. We will verify the proposed method with numerical simulations and SwellEx96 data processing.

4aUWa8. Predicting sonar performance using observations of mesoscale eddies. Harry DeFerrari (Div. of Appl. Marine Phys., RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy, Miami, FL 33149)

A predictive relationship has been observed between the location of offshore mesoscale eddies and the performance of active and passive sonar on the shallow water shelf area inside of the eddy. The passage of an eddy produces a prograde front that modifies acoustic propagation by two mechanisms. First, the density gradient serves as a conduit for offshore internal waves to propagate onto the shelf. A long-lived front can result in order of magnitude increases in potential energy of the internal wave field and corresponding increases in sound speed variability. Second, the circulation of the eddy produces a unique sound speed profile that is strongly downward refracting but has a nearly iso-velocity layer near the bottom owing to turbulent mixing. The shape of the profile closely approximates a hyperbolic cosine. Such a profile has mode group velocities that are equal for all refracted modes, thus producing strong focusing and a caustic at the depth of the source at all ranges. The experimental observations are confirmed with oceanographic and acoustic propagation models and, in turn, the models predict FOM fluctuations of as much as 15 dB for passive sonar and 24 dB for active sonar, depending on location of the eddy.

Session 4aUWb

Underwater Acoustics: Session in Honor of Leonid Brekhovskikh I

William A. Kuperman, Cochair

Scripps Inst. of Oceanography, Univ. of California, San Diego, Marine Physical Lab., MC0238, San Diego, La Jolla, CA 92093-0238

Oleg A. Godin, Cochair

NOAA, Earth System Research Lab., 325 Broadway, Boulder, CO 80305-3328

Chair's Introduction—10:15

Invited Papers

10:20

4aUWb1. Phenomenon of Leonid Maximovich Brekhovskikh as a man and a scientist. Nikolay Dubrovskiy (Andreyev Acoust. Inst., Shvernika Ul 4, Moscow 117036, Russia)

Leonid Maximovich Brekhovskikh made phenomenal contributions in acoustics: discovery of the underwater sound channel, development of the fundamental theory of wave propagation in layered media, and working out a tangent plane approximation in the wave scattering theory. Brekhovskikh contributed greatly to the organization of research teams and the dissemination of information on acoustics and oceanography through his popular books and lecturing. He also made a major breakthrough as a public figure and a statesman. He became the first Director of the Acoustics Institute at the age of 36. He served as a secretary of the Russian Academy of Sciences Branch involved in oceanography, geography, and atmospheric physics research. Brekhovskikh's achievements in science and science leadership were marked by multiple highest USSR awards and many international awards. He became an Honorary Fellow of the Acoustical Society of America and the Russian Acoustical Society. He received the Lord Raleigh Medal for the discovery that preserved its urgency for 30 years. Brekhovskikh's phenomenon is regarded here from the viewpoint of his personality as well as specific circumstances of his family and social life.

10:40

4aUWb2. Some aspects of Leonid Brekhovskikh's influence on oceanographic acoustics. W. A. Kuperman and W. Munk (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238)

Waveguide physics describes the basic features of long-range sound propagation in the ocean. Over the last half century the theory has progressed from describing ideal waveguides to more complicated layered structures to range-dependent structures to time-varying, range-dependent structures. The theme of Brekhovskikh's pioneering work was the development of robust formulations that permitted understanding basic ocean acoustics while also laying the foundation to progress to the next levels of realistic complexity. Early on, he realized that acoustic data were not consistent with known oceanography. His seminal oceanographic experiments established the pervasive presence of mesoscale phenomena, which to this day are still not fully incorporated into rigorous formulations of the forward and inverse acoustics problems. We discuss only a very small part of his work and its subsequent influence.

11:00

4aUWb3. Underwater sound propagation: 49 years with L. M. Brekhovskikh's *Waves in Layered Media*. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA, Earth System Res. Lab., 325 Broadway, Boulder, CO 80305, oleg.godin@noaa.gov)

In his first 10 years of research on wave propagation in layered media, L. M. Brekhovskikh created a theory that remains a basis for physical understanding and mathematical modeling of underwater sound propagation. Summarized in his celebrated book *Waves in Layered Media*, first published in 1957, the theory includes spectral (quasi-plane wave) representations of wave fields, normal mode theory for open waveguides, extensions of the ray and WKBJ methods, and a clear treatment of diffraction phenomena attendant to caustics, lateral waves, and reflection of wave beams and pulses. The book also charted the ways forward that have been and are followed by numerous researchers around the globe. Some of the resulting progress was documented in subsequent editions of *Waves in Layered Media* and in later books L. M. Brekhovskikh coauthored with his students. This paper will discuss diverse, groundbreaking contributions L. M. Brekhovskikh made to the wave propagation theory from the prospective offered by recent developments in underwater acoustics.

11:20

4aUWb4. L. M. Brekhovskikh's studies on nonlinear wave interaction and atmospheric sound. Konstantin Naugolnykh (Univ. of Colorado, NOAA, ESRL/Zeltaech LLD, Boulder, CO)

Nonlinear interaction of waves in a compressible fluid is an underlying factor in many geophysical effects, and L. M. Brekhovskikh made essential contributions to investigation of these phenomena. In particular, he suggested the mechanism of the infrasound generation by stormy areas in the ocean based on the nonlinear interaction of the counter-propagating sea-surface gravity

4a FRI. AM

waves. The estimates of the order of magnitude of sound intensities were made indicating that the main part of the infrasound generated by the surface waves is absorbed in the upper layers of the atmosphere, resulting in the heating of these layers. The other part of the sound energy can be trapped by the atmospheric acoustic waveguide and then returned to earth at distances of hundreds of kilometers, producing the voice of the sea.

11:40

4aUWb5. Tangent-plane approximation by L. M. Brekhovskikh and connected methods in the theory of wave scattering from rough surfaces. Alexander G. Voronovich (NOAA, Earth System Res. Lab., Physical Sci. Div., 325 Broadway, Boulder, CO 80305, alexander.voronovich@noaa.gov)

Starting from pioneering work by Rayleigh in 1907, scattering of waves from rough surfaces was restricted by the case of small Rayleigh parameter. In this case perturbation analysis describing the process of Bragg scattering applies. Apparently, smallness of the roughness is too restrictive for many applications. In 1952 L. M. Brekhovskikh suggested a tangent-plane approximation (TPA). For ideal boundary conditions it represents the first iteration of the appropriate boundary integral equation. However, for more complex situations (e.g., dielectric or solid-fluid interfaces) appropriate boundary integral equations are rather complicated and, even worse, they cannot be readily iterated. The TPA allows bypassing this step providing the answer in closed form for arbitrary boundary conditions and for scalar or vector waves in terms of the local reflection coefficient. Unfortunately, the TPA does not correctly describe the Bragg scattering. However, later it was realized that the TPA allows simple generalization, which treats both low- and high-frequency limits within single theoretical scheme. This is achieved by considering the local reflection coefficient as an operator rather than a factor. New methods going beyond the two classical ones with much wider regions of validity were developed based on this idea. Some of them will be reviewed in this talk.

FRIDAY AFTERNOON, 1 DECEMBER 2006 LANAI ROOM, 1:00 TO 4:45 P.M.

Session 4pAA

Architectural Acoustics: Measurement of Room Acoustics II

Boaz Rafaely, Cochair
Ben Gurion Univ., Electrical and Computer Engineering Dept., 84105 Beer Sheva, Israel

Hideo Miyazaki, Cochair
Yamaha Corp., Ctr. for Advanced Sound Technologies, 203 Matsunokijima, Iwata, Shizuoka 438-0192, Japan

Contributed Papers

1:00

4pAA1. Impulse response measurements based on music and speech signals. Wolfgang Ahnert, Stefan Feistel, Alexandru Miron, and Enno Finder (Ahnert Feistel Media Group, Berlin, Germany)

All known software based measurement systems, including TEF, MLSSA, SMAART, and EASERA, derive results using predetermined excitation signals like Sweep, MLS, or Noise. This work extends the range of excitations to natural signals like speech and music. In this context selected parameters like frequency range, dynamic range, and fluctuation of the signal and the signal duration are investigated in order to reach conclusions about the conditions required to obtain results comparable with standard excitation signals. Also the limitations of the standard stimuli and the proposed natural stimuli are discussed.

1:15

4pAA2. Assessment of reverberation time in halls through analysis of running music. David Conant (McKay Conant Brook Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362, dconant@mcbinc.com)

The source signal to excite a room's reverberant field sufficient for detailed measurement of reverberation time (RT60) and other measures has been the subject of considerable investigation over several decades. It is generally acknowledged that the best sources are (depending on the researcher) swept tones, MLS, MLS variations, stopped noise, cannon shots, etc. All can be characterized as highly audience unfriendly. In the

interest of obtaining useful approximations of measured midfrequency RT60 in the presence of live audiences, this paper discusses several approaches that may be fruitful while being entirely unobtrusive to the concert experience.

1:30

4pAA3. Comparison of measurement techniques for speech intelligibility. Bruce C. Olson (Olson Sound Design, 8717 Humboldt Ave N, Brooklyn Park, MN 55444, bco@olsonsound.com)

A comparison of measurement techniques for speech intelligibility between two recently released measurement systems is made. EASERA (Electronic and Acoustic System Evaluation and Response Analysis) uses a standard PC and an EASERA Gateway interface attached via Firewire. The software postprocesses a variety of stimuli in order to derive the impulse response for the room under test. This impulse response is then further processed and the results are presented to the user in both graphical and textual presentations. The Ivie Technologies IE-35 is based on a Pocket PC system and uses an external modulated noise source as stimulus to produce an intelligibility score as a single number or average of a series of measurements. This paper will explore a variety of measurements made in the same locations in a room by both systems. Results will also be shown for a variety of other acoustic measures that quantify the acoustical parameters of the room.

4pAA4. Under-balcony acoustics in concert halls: Single source versus an array of multiple sources. Youngmin Kwon and Gary W. Siebein (Architecture Technol. Res. Ctr., Univ. of Florida, 134 ARCH, P.O. Box 115702, Gainesville, FL 32611, ymkwon@ufl.edu)

The conventional measurement protocol using a single omnidirectional sound source may have limits or uncertainty in objective acoustical analysis of a performance hall. This study conducted monaural and binaural impulse response measurements with an array of 16 directional loudspeakers for quantitative acoustical assessment of specifically under-balcony area in a concert hall. The measurements were executed in a real performance hall. The measured time- and frequency-domain responses as well as the results of room acoustical parameters including binaural parameters were compared to the ones measured with a single omnidirectional source. The results were also compared to the ones taken at the main orchestra seating area. The time-domain responses showed a clear distinction particularly in early responses between single source and multiple sources. On the other hand, the magnitude of frequency response showed significantly lower at frequencies above 1 kHz than the one measured at the main area. The results of a binaural parameter, IACC, were found to be marginal between single source and multiple sources but critically different between under-balcony area and main area. Variations were also observed in the results of other room acoustical parameters when compared either between single source and multiple sources or between under-balcony area and main area.

4pAA5. Alternative metrics for the directivity of acoustic sources. Timothy W. Leishman (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

While the directivity of an acoustic source at a given frequency is thoroughly characterized by a directivity function over the angular coordinates, it may also be characterized to a lesser degree by a single-number directivity factor. The directivity index (i.e., the logarithmic version of the directivity factor) is a related figure of merit. Recent efforts to quantify the directivity of sources for architectural acoustics measurements have led to several alternatives to these values. One example is the area-weighted spatial standard deviation of radiated levels over a free-field measurement sphere. This paper presents and compares this and other directivity metrics for several types of sources, and discusses their benefits.

4pAA6. Room volume estimation from diffuse field theory. Martin Kuster and Maarten van Walstijn (Sonic Arts Res. Ctr., Queen's Univ. Belfast, BT7 1NN Belfast, Northern Ireland, m.kuster@qub.ac.uk)

Among the parameters relevant in room acoustics, the room volume is one of the most important. The general course in room acoustics research is to use the room volume in the prediction of room acoustic parameters such as reverberation time or total relative sound pressure level. Contrary to this, it has been investigated to what extent the room volume can be retrieved from a measured room impulse response. The approach followed is based on room acoustic diffuse field theory and requires correctly measured room impulse responses with the initial time delay corresponding to the source to receiver distance. A total of ten rooms of varying size and acoustic characteristics have been included. The results in three rooms were unreliable, which was explained by the particular acoustic characteristics. In the remaining rooms the results were numerically useful and consistent between different positions within the same room (relative standard deviation around 20%). The influence of source and receiver directivity is also considered.

4pAA7. *In situ* measurements for evaluating the scattering surfaces in a concert hall. Jin Yong Jeon and Shin-ichi Sato (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr)

Sound diffusion by a wall structure is one of the main concerns with respect to the sound quality of concert halls. There is a need to develop measurement and evaluation methods for determining the performance of scattering wall surfaces not only in a laboratory but also in actual halls. In this study, the acoustical measurements were conducted in a concert hall which has diffusers with ceramic cubic tiles on the side walls of the stage and the audience area. Binaural impulse responses were measured at all of the seats under two conditions, that is, with and without diffusers. The area which was affected by the diffusive wall was determined and quantified. The condition without diffusers was produced by covering them with the movable reflectors. From the binaural impulse responses, the temporal diffusion [H. Kuttruff, *Room Acoustics*, (Elsevier Science, London, 1991)], which is calculated from the autocorrelation of the impulse response, and other acoustical parameters were analyzed. From the relationship between the scattering coefficient and the acoustical parameters, sound scattering index for real halls, which represents the degree of the diffusion of a hall, was proposed.

4pAA8. Further investigations on acoustically coupled spaces using scale-model technique. Zuhre Su, Ning Xiang (Grad. Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180), and Jason E. Summers (Naval Res. Lab., Washington, DC 20024)

Recently, architectural acousticians have been increasingly interested in halls that incorporate coupled-volume systems because of their potential for creating nonexponential sound energy decay. Effects of coupling-aperture configuration and source and receiver locations on energy decay are essential aspects of acoustically coupled spaces that have not yet been extensively investigated. In order to further understand these effects on sound fields in coupled rooms, a systematic experimental study is carried out. An acoustic scale model technique is used in collecting room impulse responses of a two-room coupled system for varying aperture configurations and surface-scattering conditions. Baseline behavior is established by varying aperture area for a fixed aperture shape and analyzing relevant energy-decay parameters at different locations. Effects of aperture shape and number are systematically investigated by varying these parameters while holding coupling area fixed. Similarly, effects of receiver location are systematically investigated by varying the distance of the receiver from the coupling aperture for a fixed aperture configuration. Schroeder decay-function decompositions by Bayesian analysis reveal sensitivities to receiver location and aperture configuration across different frequency bands.

4pAA9. Virtual microphone control: A comparison of measured to created impulse responses of various microphone techniques. Daniel Valente and Jonas Braasch (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, danvprod@yahoo.com)

A method of rendering sound sources in 3-D space has been developed using virtual microphone control (ViMiC) [J. Acoust. Soc. Am. **117**, 2391]. This method has been used to create a flexible architecture for the creation and rendering of a virtual auditory environment based on microphone techniques. One of the advantages of ViMiC is the ability to simulate coincident, near-coincident, and spaced microphone recording techniques. This allows the user active spatial control over the recorded environment and the ability to shape the final rendering based on his or her specific auditory needs. In order to determine the accuracy of simulating the virtual microphone techniques, measurements of several acoustic

4p FRI. PM

spaces in Troy, NY will be compared to the measurements of generated impulse responses of the same modeled spaces within the ViMiC environment. The data from the measured impulse responses will be used to adapt the ViMiC system in order to create a more realistic auditory rendering. Moreover, the ViMiC system can be improved for use as an educational tool for teaching recording engineers to hear the subtle differences between various microphone techniques.

3:30

4pAA10. The estimation of the room acoustic characteristic using the acoustic intensity method. Yong Tang, Hideo Shibayama, and Takumi Yosida (Dept. of Commun. Eng., Shibaura Inst. of Technol., 3-7-5 Toyosu Koutou-ku, Tokyo, 135-8548 Japan, m603101@sic.shibaura-it.ac.jp)

When a sound radiates in rooms, a lot of reflection sounds are generated. From estimation of the direction where the room reflection sound comes from, we can understand the diffusion situation in the room acoustic field. By using the acoustic intensity method, we can measure the strength and the direction of the sound. In this paper, we estimate the direction of the reflection sound in the time-space by the acoustic intensity method and show the acoustic characteristic of the room.

3:45

4pAA11. Binaural simulation in an enclosure using the phased beam tracing. Cheol-Ho Jeong and Jeong-Guon Ih (NOVIC, Dept. of Mech. Eng., KAIST, Sci. Town, Daejeon 305-701, Korea, chjeong@kaist.ac.kr)

Binaural simulation in an enclosed space is important in the subjective evaluation of the enclosure acoustics in the design or refinement stage. A time domain scheme using the geometrical acoustics technique has been usually used in the binaural processing. However, one can calculate a pressure impulse response by using the phased beam-tracing method, which incorporates the phase information in the beam tracing process. Such phased method employs reflection coefficient and wave number, whereas the conventional method uses absorption coefficient and air attenuation factor. Impulse response can be obtained by the inverse Fourier transformation of the frequency domain result. This feature facilitates the binaural simulation because the convolution with the HRTF can be accomplished by a simple multiplication in frequency domain. Convolutions were conducted for all reflections one by one, and the convolved transfer functions were summed into one transfer function. Consequently binaural room impulse responses at receivers' ear positions can be simulated. The measured binaural room impulse responses in the conference room were compared with the predicted results for octave bands of 125 Hz to 4 kHz. A good agreement with measurement was found, especially in the early part of impulse responses. [Work supported by BK21.]

4:00

4pAA12. Visualization methods of direct and early reflection sounds in small enclosures. Chiaki Koga, Akira Omoto (Omoto Lab., Dept. of Acoust. Design, Faculty of Design, Kyushu Univ., Shiobaru 4-9-1, Minami, Fukuoka 815-8540, Japan), Atsuro Ikeda, Masataka Nakahara (SONA Corp., Nakno-ku, Tokyo, 164-0013, Japan), Natsu Tanaka, and Hiroshi Nakagawa (Nittobo Acoust. Eng. Co., Ltd. Sumida-ku, Tokyo 130-0021, Japan)

Many parameters exist for evaluating large sound fields such as concert halls. However, it is difficult to apply those parameters for evaluation of a small room such as a recording studio because of their different sound fields. Widely useful common parameters have not been established.

Moreover, early reflections are important in small rooms for determining spatial acoustic impressions. Therefore, various methods that visualize spatial acoustic information obtained by early reflection in rooms are proposed. For this study, sound fields (a music studio and a filmmaking studio) were measured using three kinds of different techniques: instantaneous intensity, mean intensity, and a sphere-baffled microphone array. This report compares the information of sound source directions obtained using these methods. Results show that every method can estimate the position of sound sources and important reflections with high accuracy. In the future, we shall propose a method that visualizes spatial acoustic information more precisely by combining the methods and establishing acoustic parameters that are available for evaluating and designing small rooms.

4:15

4pAA13. Acoustic evaluation of worship spaces in the city of Curitiba, Brazil. Cristiane Pulsides, David Q. de Sant'Ana, Samuel Ansay (LAAICA/UFPR, Bloco 4 sala PG-05 81531-990 Curitiba, PR, Brasil, pulsides@gmail.com), Paulo Henrique T. Zannin, and Suzana Damico (LAAICA/UFPR, 81531-990 Curitiba, PR, Brasil)

This article searches acoustic parameters in religious buildings located in the city of Curitiba intending to study its behavior in this kind of facilities. The temples were analyzed according to type of ceremony, architectonic style, and construction date. The research was made through the impulsive response integration method for three energetic parameters: (1) reverberation time (RT); (2) clarity (C80); and (3) definition (D50) according recommendations of the ISO/3382:1997 Standard. Performed in between were six and eight impulsive responses in each room using sweep signals and omnidirectional microphones. The results were than compared with referential values already existing [W. Fasold and E. Veres, *Schallschutz + Raumakustik in der Praxis*, 136 (1998)] for acoustic characterizations. It is possible to observe in the measurements the direct connection between reverberation time and the parameters clarity or definition. Moreover, it is possible also to observe the influence of the geometric ratios and architectural elements of the rooms, getting itself for equivalent volumes and rays of removal of the source, different levels of definition.

4:30

4pAA14. A consideration of the measurement time interval for obtaining a reliable equivalent level of noise from expressway. Mitsunobu Maruyama (Salesian Polytechnic, Oyamagaoka 4-6-8, Machida, Tokyo 194-0215, Japan) and Toshio Sone (Akita Prepectural Univ., Honjo, Akita 015-0055, Japan)

The level of road traffic noise $LA_{eq,T}$ greatly depends on the maximum level during the measurement time interval tau, and the maximum level often appears at the moment when two consecutive heavy vehicles pass through the point adjacent to the observation point. A mathematical model is proposed for simulating the variation in traffic noise, especially from the point of heavy vehicles with passing. The mean time interval between a pair of two consecutive heavy vehicles with the minimum allowable distance is obtained from time-series data and the mean recurrence time h_{ij} which can be calculated from the transition matrix $P[p_{ij}]$. The comparative study is made among the numbers of heavy vehicles from 25 to 300 [vehicles/hour] in traffic flow and the observation distances of 40 to 200 m from the road. The result shows that the measurement time interval required for the acquisition of reliable data is three to four times as long as tau or h_{ij}.

Session 4pABa

Animal Bioacoustics: Marine Mammal Acoustics II

David K. Mellinger, Chair

Oregon State Univ., Hatfield Marine Science Ctr., Newport, OR 97365

Contributed Papers

1:15

4pABa1. Great ears: Functional comparisons of land and marine leviathan ears. D. R. Ketten (Harvard Med. School, Boston, MA; Woods Hole Oceanograph. Inst., Woods Hole, MA), J. Arruda, S. Cramer, M. Yamato (Woods Hole Oceanograph. Inst., Woods Hole, MA), J. O'Malley (Massachusetts Eye and Ear Infirmary, Boston, MA), D. Manoussaki (Vanderbilt Univ., Nashville, TN), E. K. Dimitriadis (NIH/NIDCD, Bethesda, MD), J. Shoshani (Univ. of Asmara, Asmara, Eritrea), and J. Meng (American Museum of Natural History, New York, NY)

Elephants and baleen whales are massive creatures that respond to exceptionally low frequency signals. Although we have many elephant and whale vocalization recordings, little is known about their hearing. Playback experiments suggest hearing in both proboscideans and mysticetes is tuned similarly to low or even infrasonic signals. This raises several interesting issues. First, they emit and perceive signals in two media, air and water, with radically different physical acoustic properties: 4.5-fold differences in sound speed, three-fold magnitude difference in acoustic impedance, and, for common percepts, whales must accommodate 60-fold acoustic pressures. Also, a commonly held tenet is that upper hearing limit is inversely correlated with body mass, implying there should be virtually no whale-elephant hearing overlap given body mass differences. This study analyzed how inner ears in these groups are structured and specialized for low-frequency hearing. Computerized tomography and celloidin histology sections were analyzed in six baleen whale ($n = 15$) and two elephant species ($n = 7$). The data show mysticetes have a substantially greater hearing range than elephants but that coiling and apical cochlear structures are similar, suggesting common mechanical underpinnings for LF hearing, including cochlear radii consistent with the Whispering Gallery propagation effect. [Work supported by ONR, NIH, WHOI OLI, Seaver Foundation.]

1:30

4pABa2. Social context of the behavior and vocalizations of the gray whale *Eschrichtius robustus*. Sarah M. Rohrkasse (School for Field Studies, Ctr. for Coastal Studies, Apartado Postal 15, Puerto San Carlos, BCS, CP 23740 Mexico, sarro101@hotmail.com) and Margaret M. Meserve (Guilford College, Greensboro, NC 27410)

Sound production and surface behavior of the gray whale were investigated at Bahia Magdalena, Mexico to determine if vocalizations have behavioral correlations or are used in specific social contexts. Fifteen-minute sessions of behavioral observations and acoustic recordings of gray whales in various social contexts were collected from February to April 2006 ($n = 30$). Analysis of sound production included proportional use of different call types and acoustic variables of each sound type. Preliminary acoustic analysis found no correlation with social contexts or behaviors, but proportional use of different vocalizations is similar to past studies in Baja [Dahlheim *et al*, *The Gray Whale*, pp. 511–541 (1984), F. J. Ollervides, dissertation, Texas A&M University (2001)]. Initial results indicate significant differences in frequencies of high surface behaviors ($p = 0.0477$) of groups that include mother-calf pairs. As analysis continues, possible correlations between social context and use of sounds could allow for acoustics to be an indicator of group composition, seasonal movements, and social patterns and to help determine the functions of sounds. [Work supported by SFS and NFWF.]

1:45

4pABa3. Ambient noise and gray whale *Eschrichtius robustus* behavior. Francisco Ollervides, Kristin Kuester, Hannah Plekon, Sarah Rohrkasse (School for Field Studies—Ctr. for Coastal Studies, Apartado Postal 15, Puerto San Carlos, BCS, CP 23740 Mexico, follervides@hotmail.com),Kristin Kuester (Univ. of Wisconsin−Madison, Madison, WI 53706), HannahPlekon (Davidson College, Davidson, NC), andSarahRohrkasse (Texas A and M Univ., College Station, TX 77843)

Between 14 February and 13, April 2006, we conducted 31 recording sessions of ambient noise and behavioral sampling of gray whales within Magdalena Bay, Mexico. This breeding lagoon does not have the same Marine Protected Area status compared to the other breeding lagoons of San Ignacio and Guerrero Negro in the Mexican Pacific coast. Poorly monitored guidelines and increasing boat traffic from whale−watching tourism in this area have the potential to affect the surface behavior of these animals and increase average ambient noise levels. Relative ambient noise levelswererecordedandcomparedtoapreviousstudy[Ollervides,2001]todetermine similarities or differences in the 5−year interval between both data sets. Although results are not comparable in decibel levels, probably due to equipment calibration problems, there was a significant difference between the different regions of the bay Kruskal−Wallis ($p = 0.0067$). Activity levels ranged from 0.005–0.196 behaviors/whale/minute. Ambient noise levels ranged from 35.70–64.32 dB Re: 1 Pa. No correlation was found between the ambient noise levels in the bay and the activity level of gray whales (correlation value=0.0126; log correlation value=0.172). Further acoustic processing is currently underway.

2:00

4pABa4. Look who's talking; social communication in migrating humpback whales. Rebecca A. Dunlop, Michael J. Noad (School of Veterinary Sci., Univ. of Queensland, St. Lucia, Qld 4072, Australia. r.dunlop@uq.edu.au), Douglas H. Cato (Defence Sci. and Tech Org., Pyrmont, NSW 2009, Australia), and Dale Stokes (Scripps Inst. of Oceanogr., La Jolla, CA 92037)

A neglected area of humpback acoustics concerns nonsong vocalizations and surface behaviors known collectively as social sounds. This study describes a portion of the nonsong vocal repertoire and explores the social relevance of individual sound types. A total of 622 different sounds were catalogued and measured from whales migrating along the east coast of Australia. Aural and spectral categorization found 35 different sound types, and discriminate functions supported 33 of these. Vocalizations were analyzed from 60 pods that were tracked visually from land and acoustically using a static hydrophone array. Nonsong vocalizations occurred in all pod compositions: lone whales, adult pairs, mother/calf pairs, mother/calf/escorts, and multiple-adult pods. Thwops and wops were likely to be sex-differentiated calls with wops from females and thwops from males. Sounds similar to song-units were almost all from joining pods and yaps were only heard in splitting pods. Other low-frequency calls (less than 60 Hz) were thought to be within-pod contact calls. Higher-frequency cries (fundamental 450–700 Hz) and other calls (above 700 Hz) and presumed underwater blows were heard more frequently in joining

3265   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3265**

4p FRI. PM

pods displaying agonistic behaviors. This work demonstrates that humpbacks produce a great range of contextually different communication signals. [Work supported by ONR and DSTO.]

2:15

4pABa5. Seasonal ambient noise levels and impacts on communication in the North Atlantic right whale. Susan E. Parks, Christopher W. Clark, Kathryn A. Cortopassi, and Dimitri Ponirakis (Bioacoustics Res. Program, Cornell Univ., 159 Sapsucker Woods Rd., Ithaca, NY 14850, sep6@cornell.edu)

The North Atlantic right whale is a highly endangered species of baleen whale. Acoustic communication plays an important role in the social behavior of these whales. Right whales are found in coastal waters along the east coast of the United States, an area characterized by high levels of human activity. Most of these activities generate noise that is propagated into the coastal marine environment. The goals of this project are to characterize the noise, both natural and anthropogenic, in right whale habitat areas to determine what levels of noise the whales are regularly exposed to, and whether the acoustic behavior of right whales changes in response to increased noise. Continuous recordings were made from autonomous bottom-mounted recorders in three major habitat areas in 2004 and 2005; Cape Cod Bay (December–May), Great South Channel (May), and the Bay of Fundy, Canada (August) to passively detect right whales by recording their vocalizations. Here, we describe the ambient noise levels in these recordings to describe the daily acoustic environment of right whales, how noise varied over diel, weekly, and seasonal time scales, and whether noise levels correlated with any observed changes in acoustic behavior of the whales.

2:30

4pABa6. Blue whale calling in Australian waters. Robert D. McCauley, Chandra P. Salgado Kent (Curtin Univ. of Technol., G.P.O. Box U 1987, Perth 6845, Australia), Christopher L.K. Burton (Western Whale Res. Hillarys 6923, WA Australia), and Curt Jenner (Ctr. for Whale Res. (WA Inc.), Fremantle WA, 6959 Australia)

Calling from the Antarctic true blue whale (*Balaenoptera musculus intermedia*) and the tropical subspecies (brevicauda, or pygmy blue) have been recorded across southern Australia with the pygmy blue calls also recorded along the Western Australian (WA) coast. The subspecies have a believed common downsweep and markedly different longer, tonal calls. The frequency of most energy in the tonal calls is offset between the subspecies suggesting sound-space partitioning. The pygmy blue three-part tonal call is typically 120 s long repeated every 200 s, has several variants, and includes a complex two-source component. The nature of the pygmy blue call allows counts of instantaneous calling individuals, giving relative abundance. These estimates in the Perth Canyon, a localized seasonal feeding area, show patterns in usage of space and through time within and between seasons, such as the sudden departure of animals at a season end, which varies by approximately 2 weeks between years. Sea noise records along the WA coast indicate south-traveling animals arrive midway along the coast in October to November, animals fan out across southern Australian over December through May, then move north in the Austral winter. We have begun converting abundance estimates from relative to absolute for pygmy blue calling rates.

2:45

4pABa7. Acoustical monitoring of finback whale movements on the New Jersey Shelf. Altan Turgut (Naval Res. Lab., Acoust. Div., Washington, DC 20375) and Christopher Lefler (Univ. of California Santa Barbara, Santa Barbara, CA 93106)

Acoustical monitoring of finback whales is performed by using a data set collected over a 3-week period in December of 2003 on the New Jersey Shelf. One-second-duration 20-Hz signals of finback whales were recorded on three vertical line arrays (VLAs) and a bottomed horizontal line array (HLA). One-second-duration pulses are separated by about 10 s and there is an approximately 2-min-long silent period between 10- to 18-min-long pulse trains. A 30- to 60-min silent period after 5 to 10 pulse trains is also common. Modal analysis of individual pulses indicated that most signals contained two acoustic modes. Arrival-time and group-speed differences of these modes are used for remote acoustic ranging. These modal characteristics are also exploited in a broadband matched-field algorithm for depth discrimination. Bearing estimation of individual whales is obtained by performing horizontal beamforming on the HLA data. Range estimation results are verified by time-of-flight triangulation using single hydrophone data from each VLA location. Acoustic monitoring results indicated that most finback whales traveled near the shelf break front where food might be abundant. Relations between silent periods and acoustic range/depth monitoring results are also investigated. [This work was supported by the ONR.]

3:00–3:15   Break

3:15

4pABa8. Analysis of melon-headed whale aggregation in Hanalei Bay, July 2004. David M. Fromm (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5350), Joseph R. Mobley, Jr. (Univ. of Hawaii at M_noa, Honolulu, HI 96822), Stephen W. Martin (Space and Naval Warfare Systems Ctr. San Diego, San Diego, CA 92152-5001), and Paul E. Nachtigall (Univ. of Hawaii at M_noa, Kailua, HI 96734)

On 3 July 2004, an aggregation of ca. 150–200 melon-headed whales (*Peponocephala electra*) appeared in the shallow waters of Hanalei Bay, Kauai and congregated there for over 27 h. Preceding the whales' appearance and partially coincident with their time in the Bay, midrange (3.5–5 kHz) tactical sonars were intermittently deployed during the Rim of the Pacific 2004 (RIMPAC) joint military exercises being conducted in waters near Kauai by the U.S., Japan, and Australia Navies. An NOAA report (Southall *et al.*, 2006) attributed the active sonar usage as a plausible, if not likely, contributing factor. A detailed timeline and reconstruction of the RIMPAC activities is presented showing the worst-case estimates of the sonar sound levels in the waters surrounding Kauai. A re-examination of available evidence combined with a new report of a simultaneous and similar aggregation in Sasanhaya Bay, Rota, Commonwealth of the Northern Mariana Islands, brings the plausibility conclusion into question. [This work was sponsored by multiple sources. D. Fromm and S. Martin conducted acoustic analyses with funds provided by the U.S. Pacific Fleet. J. Mobley received funding from the U.S. Geological Survey. P. Nachtigall is sponsored by the Office of Naval Research for marine mammal audiometric studies.]

3:30

4pABa9. Midfrequency sound propagation in beaked whale environments. Eryn M. Wezensky, Thomas R. Stottlemyer, Glenn H. Mitchell (Naval Undersea Warfare Ctr., Newport Div., Newport, RI 02841), and Colin D. MacLeod (Univ. of Aberdeen, Aberdeen, U.K.)

Recent mass strandings of beaked whales (*Ziphiidae, Cetacea*) coinciding with the use of midfrequency range (1–10 kHz) active sonar have caused speculation about the potentially adverse effects of these sound sources. Particular questions of the research and regulatory communities concern whether beaked whale sensitivity to midfrequency sound exposure is influenced by oceanographic characteristics present at the time of the mass stranding events. This study investigated the interaction between beaked whale habitat characteristics and the nature of a midfrequency signal by analyzing the oceanographic factors affecting underwater acoustic propagation. Three types of model sites were selected from five specific geographical locations where beaked whales have been regularly recorded or where a mass stranding event has been reported. A ray-trace acoustic propagation model was used to generate transmission loss for a 3-kHz signal over a representative 60-km transect at each locality. Model outputs visually demonstrated how the combination of site/event-specific oceanographic characteristics affects the sound propagation of a moving source. A parametric sensitivity comparison and statistical analysis were conducted to identify influential factors between environmental parameters, source depth, and the resulting transmission loss. Major findings of this study as well as future research direction are discussed. [Research supported by NAVSEA.]

3:45

4pABa10. Examination and evaluation of the effects of fast rise-time signals on aquatic animals. Michael Stocker (Seaflow, Inc., 1062 Ft. Cronkhite, Sausalito, CA 94965)

Increasingly human enterprise is subjecting the ocean environment to acoustic signals to which marine animals are not biologically adapted. This is evidenced by a marked rise in marine mammal strandings, as well as hearing and other physiological damage to fish and other marine organisms as a result of, or coincident to, human-generated noise events. Determining phonotoxic thresholds of marine organisms is complicated by the fact that various marine animals are adapted to sense either pressure gradient or particle motion acoustic energy, or some combination or gradient between the two. This has been addressed to some degree by exposure metrics that consider either net or accumulated acoustical flux densities from various noise sources. This paper examines the role and effects of signal rise time both in terms of physiological impulse response of the exposed organisms, as well as broadband saturation flux densities of fast rise-time signals on animal sense organs. Case studies from the literature will be presented to demonstrate the effects of fast rise time signals on fish. Acoustical signals with high crest factors and fast rise-time components will be compared to signals with dominantly sinusoidal components to illustrate the perceptual effects of these signals on human hearing.

4:00

4pABa11. Noise separation of underwater acoustic vocalization using auditory filter bank and Poisson rate estimation. Owen P. Kenny and Craig R. McPherson (Dept. of Elec. and Comput. Eng., James Cook Univ., Douglas 4811, Queensland, Australia)

Formant vocalization tracking has been achieved using a mammalian periphery model and a Poisson rate estimator. This approach used a set of linear bandpass filters to simulate the mechanical displacement of the basilar membrane. The auditory model simulated neural firing by producing a spike at the positive going zero crossing for each filter output. Poisson

intensity of the neural firing rate is controlled by the dominant frequency components of the signal present in the filter. This approach is extended by incorporating neural synchronization information to separate the formant structure from that of noise. The filter structure is designed to overlap the frequency range of adjacent filters. The presence of a formant structure in adjacent filters controls the interspike intervals of neural firing for both filters, which results in the neural firing from both filters being synchronized. If a noise-only component is present in either filter, then the spiking outputs from the adjacent filters are unsynchronized. Experimental results have shown that incorporating neural synchronization information between adjacent filters has enabled separation of signal components from noise. This technique enables easier signal and noise separation than allowed by traditional methods.

4:15

4pABa12. Using vocalizations of Antarctic seals to determine pupping habitats. T. L. Rogers, C. J. Hogg, M. B. Ciaglia (Australian Marine Mammal Res. Ctr., Zoological Parks Board of NSW/Faculty of Veterinary Sci., Univ. of Sydney, Mosman Australia), and D. H. Cato (Defence Sci. & Technol. Organisation, Pyrmont, Australia)

The Ross and Leopard seal use the floes of the Antarctic pack ice to whelp and raise their pups, But both species are rarely seen in summer throughout the pack ice. We now realize that this is because they are under the water "calling" during the austral summer as part of their breeding display, and so their presence is underestimated in traditional visual surveys. The period of "calling" overlaps with the time that females give birth, so their vocalizations can be used to determine seal distributions during this time. Acoustic recordings were made using sonobuoys deployed during ship based surveys in the pack ice and analyzed to determine the seal distributions. This was used to predict habitat preference of seals by relating their distributions to remotely sensed indices: ice cover, ice floe type, ice thickness, distance to ice edge, distance to shelf break, distance to land, sea surface temperature, and chlorophyll a.

FRIDAY AFTERNOON, 1 DECEMBER 2006

KOHALA/KONA ROOM, 4:30 TO 5:15 P.M.

Session 4pABb

Animal Bioacoustics: Avian Acoustics

Ann E. Bowles, Chair
Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Contributed Papers

4:30

4pABb1. Effective area of acoustic lure surveys for Mexican spotted owls (*Strix occidentalis lucida*). Samuel L. Denes, Ann E. Bowles (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109, sdenes@hswri.org), Kenneth Plotkin, Chris Hobbs (Wyle Labs., Arlington, VA 22202), John Kern (Kern Statistical Services, Sauk Rapids, MN 56379), and Elizabeth Pruitt (GeoMarine, Inc., Hampton, VA 23666)

During acoustic lure surveys for birds, topography and ambient noise are likely to be important determinants of detectability. Examinations of propagation were conducted for acoustic lures (human-made calls) and owl responses recorded during acoustic surveys for Mexican spotted owls in the Gila National Forest (2005). Lure surveys were designed based on

formal agency protocols, which assumed a 0.43-km detection range under typical conditions. A total of 558 points was called over a heavily forested, topographically complex 20×24-km area. Real-time measurements of owl calls and lures were made with a calibrated recording system. Ambient noise was collected using an array of 39 Larson-Davis 820 and 824 sound-level meters. The NMSIM (Wyle Laboratories) single-event propagation simulator was used to model propagation of both owl and human calls. The resulting model of survey effort was compared with a simple two-dimensional statistical model. Probability of detecting owls did not fit the expectations of the agency protocol, suggesting that acoustic propagation should be considered during owl surveys. [Work supported by U.S. Air Force ACC/CEVP; USFWS Permit No. TE024429]

4:45

4pABb2. Automated localization of antbirds and their interactions in a Mexican rainforest. Alexander N. G. Kirschel, Travis C. Collier, Kung Yao, and Charles E. Taylor (Univ. of California, Los Angeles, 621 Charles E. Young Dr. South, Los Angeles, CA 90095)

Tropical rainforests contain diverse avian communities incorporating species that compete vocally to propagate their signals to intended receivers. In order to effectively communicate with birds of the same species, birds need to organize their song performance temporally and spatially. An automated identification and localization system can provide information on the spatial and temporal arrangement of songs. Acoustic sensor arrays were tested for the ability to localize the source of songs of antbirds recorded in a Mexican rainforest. Pilot studies with a five-node array arranged in a rough circle with a 20-m diameter located the song of Dusky Antbird (*Cercomacra tyrannina*) with an error of 73 cm and Mexican Antthrush (*Formicarius moniliger*) with an error of 65 cm from the location of a source loudspeaker within the array. An additional source 21 m outside was also localized. Results will be presented for experiments and recordings of individuals at the Mexican rainforest site in October 2006. Locations of birds of the same and different species during vocal performance will provide a greater understanding of how individuals interact spatially with each other based on their vocal performance, from which the role of song in ecological interactions can be inferred.

5:00

4pABb3. Nonintrusive acoustic identification of hermit thrush (*Catharus guttatus*) individuals. Dennis F. Jones (Defence R&D Canada—Atlantic, P.O. Box 1012, Dartmouth, NS, Canada B2Y 3Z7, dennis.jones@drdc-rddc.gc.ca)

From mid-April well into the summer, the secretive hermit thrush (*Catharus guttatus*) can be heard singing throughout the woodlands of Nova Scotia. Its song is distinctive, beginning with a clear introductory note followed by a flurry of flutelike body notes, often cascading and reverberant in character. Despite this fine display of avian virtuosity, few studies have been reported that probe the differences between the calls, songs, and repertoires of individuals. From April 2003 to May 2006, over 3000 songs from several birds were recorded using digital video cameras at study sites in and around the city of Halifax, Nova Scotia. The only birds recorded were those in close proximity to roads and trails. None of the birds were marked, banded, or deliberately disturbed in any way. Although the study birds remained hidden from view most of the time, in the few instances where the birds perched in the open, their behaviors while singing were captured on videotape. All of the birds were readily distinguishable from each other as no two individuals had a single song in common. The most significant finding was that individuals could be reidentified acoustically after 1 week, 3 months, and 1 year had elapsed.

FRIDAY AFTERNOON, 1 DECEMBER 2006　　　　　　　　　　KAHUKU ROOM, 1:00 TO 4:50 P.M.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration and Signal Processing in Acoustics: Elastic Imaging

Peter J. Kaczkowski, Cochair

Univ. of Washington, Applied Physics Lab., 1013 NE 40th Street, Seattle, WA 98105-6698

Tsuyoshi Shiina, Cochair

Univ. of Tsukuba, Graduate School of Systems and Information Engineering, 1-1-1 Tennodai, Tsukuba 305-8573, Japan

Invited Papers

1:00

4pBB1. Present and future of elasticity imaging technology. Tsuyoshi Shiina (Grad. School of Systems and Information Eng., Univ. of Tsukuba, 1-1-1 Tennodai Tsukuba, Japan) and Ei Ueno (Univ. of Tsukuba, Tsukuba, Japan)

Elastic properties of tissues are expected to provide us novel diagnostic information since they are based on tissue characteristics and sensitively reflect its pathological state. So far, various techniques for tissue elasticity imaging have been proposed. However, it was not so easy to satisfy real-time operation and freehand manipulation of probe, which was required for practical equipment. To satisfy these conditions, we developed the combined autocorrelation method (CAM) and recently manufactured a commercial ultrasound scanner, for real-time tissue elasticity imaging by implementing the CAM algorithm, By slightly compressing or relaxing the body through freehand operation, the strain images are obtained with real-time and superimposed on B-mode images with a translucent color scale. In addition, we proposed elasticity scores of malignancy by categorizing patterns of elasticity images of breast tumors into five classes from malignant to benign. As a result of diagnosis based on the elasticity score, it was revealed that even nonexperts could attain precise diagnosis of breast cancer based on elasticity score as well as experts since the criterion on elasticity score is much simpler than conventional B-mode images. Finally, some prospects for the next stages of elasticity imaging technology well be surveyed.

1:20

4pBB2. Real-time tissue elasticity system—Development and clinical application. Takeshi Matsumura, Tsuyoshi Mitake (Hitachi Medical Corp. 2-1, Toyofuta, Hashiwa-Shi, Chiba-Ken, Japan), Tsuyishi Tsuyishi, Makoto Yamakawa, Ei Ueno (Tsukuba Univ.), Nobuhiro Fukunari (Shouwa Univ.), and Kumi Tanaka (Nippon Medical Univ.)

The progress of recent semiconductor technology has a remarkable thing. Thanks to progress of this semiconductor technology, the ultrasound scanner in medicine could come to hold enormousness computing power and has come to realize various complicated processing. At the same time, hardness of human tissue which, as you know, is used by palpation is already the information that is important in a diagnosis. But, we think that it does not have enough objectivity. To increase objectivity by visualizing hardness of

tissue, we adopted ECAM (extended combined autocorrelation method), which was developed by Professor Shiina at Tsukuba University in Japan, and succeeded in developing the commercial ultrasound scanner, which could display a strain image in real time. From a clinical point of view, in the breast region, mammography examination is effective in a diagnosis, but a judgment of permeation degree is not superior in ultrasound image. And in a thyroid gland region, we begin to get experience with availability from a diagnosis of papillary cancer and follicular cancer. So, we would like to have the presentation about the development of a strain imaging function and some of our clinical experiences by using the developed system.

1:40

4pBB3. Elasticity of perfused tissue. Kirk W. Beach (Dept. of Surgery, Univ. of Washington, Box 356410, Seattle, WA 98195-6410), Barbrina Dunmire, and John C. Kucewicz (Univ. of Washington, Seattle, WA 98105-6698)

Elastic imaging intends to measure Young's modulus (tissue stiffness) or bulk modulus (tissue compressibility) of tissue subjected to an applied strain of several percent. Underlying elastic imaging is the assumption of a linear stress/strain relationship without hysteresis or other time-dependent behavior. Perfused tissue is a composite material comprised of a solid matrix of cells, fibers, interstitial fluid (occupying up to 50% of the tissue volume and varying slowly with time), arterioles (pulsating high-pressure spaces that occupy 0.1% of the tissue volume), capillaries, and venules (low-pressure spaces that occupy up to 3% of the tissue volume varying with respiration). This talk will speculate on the nonlinear, nonstationary stress/strain relationships expected from dependent tissues (legs), pressurized tissues (breast tumors), and other living, perfused tissues. The pressure versus strain curve from each tissue voxel allows the measurement of arteriolar and venular volumes and pressures, and interstitial pressure within the tissues. These volumes and pressures may be key to classifying pathologies.

2:00

4pBB4. New developments in transient elastography. Mathias Fink, Mickael Tanter, Ralph Sinkus, and Gabriel Montaldo (LOA, ESPCI, 10 rue Vauquelin, 75005, Paris, France)

An ultra-high-rate ultrasonic scanner has been developed that can give 5000 ultrasonic images per second of the body. With such a high frame rate, the propagation of transient shear waves can be followed, and from the spatio-temporal evolution of the displacement fields, various inversion algorithms allow us to recover the shear modulus map. A discussion on the various inversion algorithms will be presented. In order to obtain unbiased shear elasticity map, different configurations of shear sources induced by radiation pressure of focused transducer arrays are used. Both 2-D and 3-D imaging can obtained with this technique. *In vitro* and *in vivo* results on breast will be presented that demonstrate the interest of elasticity imaging with transient elastography.

Contributed Papers

2:20

4pBB5. Spectral characteristics of breast vibro-acoustography images. Azra Alizad, Dana H. Whaley, Mathew Urban, Randall R. Kinnick, James F. Greenleaf, and Mostafa Fatemi (Mayo Clinic College of Medicine, Rochester, MN 55905 aza@mayo.edu)

Vibro-acoustography image is a function of the dynamic characteristics of the object at the vibration (difference) frequency (df). The dynamic characteristic of tissue is closely related to pathology. Therefore, it is important to evaluate image features versus df. Here, the influence of df on breast vibro-acoustography images is studied by scanning human breast at various df values ranging from 20 to 90 kHz. The subjects were chosen from a group of volunteers with different breast abnormalities. Images were compared subjectively to study image features and the appearances of breast lesions versus df. It is demonstrated that having a collection of images of the same tissue at different df values generally provides a better perception of the tissue structure and improves lesion identification. In most cases, higher df resulted in a higher signal-to-noise ratio and thus a higher image quality. Finally, a frequency-compounded images was obtained by calculating the weighted sum of images at different df values. It is demonstrated that image compounding normally improves visualization of breast tissue and abnormalities. [Work supported by NIH Grant EB-00535 and Grant BCTR0504550 from the Susan G. Komen Breast Cancer Foundation. Disclosure: Parts of the techniques used here are patented by MF and JFG.]

2:35

4pBB6. Tissue pulsatility imaging: Ultrasonic measurement of strain due to perfusion. John C. Kucewicz, Barbrina Dunmire, Lingyun Huang, Marla Paun (Univ. of Washington Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105-6698), and Kirk W. Beach (Univ. of Washington, Seattle, WA 98195-6410)

Over each cardiac cycle perfused tissues expand and relax by a fraction of a percent as blood rapidly accumulates in the arterial vasculature during systole and then slowly drains through the venous vasculature during diastole. Tissue pulsatility imaging (TPI) is a variation on ultrasonic tissue strain imaging that estimates tissue perfusion from this natural, cyclic tissue expansion and relaxation. TPI is derived in principle from plethysmography, a century-old technology for measuring gross tissue volume change from a whole limb or other isolatable body part. With TPI, the plethysmographic signal is measured from hundreds or thousands of sample volumes within an ultrasound image plane to characterize the local perfusion throughout a body part. TPI measures tissue strain over the cardiac cycle and parametrizes the signal in terms of its amplitude and shape. The amplitude of the strain waveform is correlated with perfusion, and the shape of the waveform is correlated with vascular resistance. Results will be presented from the leg showing the change in the TPI signals as the muscles recover from exercise, from breast tumors, and from the brain as blood flow changes in response to visual stimulation. [Work supported in part by NIH 1-R01EB002198-01 and NIH N01-CO-07118.]

2:50

4pBB7. Using human body shear wave noise for passive elastography.
Karim G. Sabra, Stephane Conti, Philippe Roux, and William A. Kuperman (Scripps Inst. of Ocean., Univ. of California—San Diego, 9500 Gilman Dr., San Diego, CA 92093-0238)

An elastography imaging technique based on passive measurement of shear wave ambient noise generated in the human body (e.g., due to the heart, muscles twitches, and blood flow system) has been developed. This technique merges two recent research developments in medical imaging and physics: (1) recent work on the efficacy of elastographic imaging demonstrating that shear waves are excellent candidates to image tissue elasticity in the human body and (2) theory and experimental verification in ultrasonics, underwater acoustics, and seismology of the concept of extracting coherent Green's function from random noise cross correlations. These results provide a means for coherent passive imaging using only the human body noise field, without the use of external active sources. Coherent arrivals of the cross correlations of recordings of human body noise in the frequency band 2–50 Hz using skin-mounted accelerometers allows us to estimate the local shear velocity of the tissues. The coherent arrivals emerge from a correlation process that accumulates contributions over time from noise sources whose propagation paths pass through both sensors. The application of this passive elastography technique for constructing biomechanical models of *in vivo* muscles' properties will be discussed.

3:05−3:20   Break

3:20

4pBB8. Dynamic radiation force of acoustic waves on solid elastic spheres.
Glauber T. Silva (Instituto de Computação, Universidade Federal de Algoas, Maceió, AL, 57072-970, Brazil)

The present study concerns the dynamic radiation force on solid elastic spheres exerted by a plane wave with two frequencies (bichromatic wave) considering the nonlinearity of the fluid. Our approach is based on solving the wave scattering for the sphere in the quasilinear approximation within the preshock wave range. The dynamic radiation force is then obtained by integrating the component of the momentum flux tensor at the difference of the primary frequencies over the boundary of the sphere. Effects of the fluid nonlinearity play a major role in dynamic radiation force, leading it to a regime of parametric amplification. The developed theory is used to calculate the dynamic radiation force on three different solid spheres (aluminum, silver, and tungsten). The obtained spectrum of dynamic radiation force presents resonances with larger amplitude and better shape than those exhibited in static radiation force. Applications of the results to some elasticity imaging techniques based on dynamic radiation force will be presented.

3:35

4pBB9. Ultrasonic measurement of displacement distribution inside an object caused by dual acoustic radiation force for evaluation of muscular relax property due to acupuncture therapy.
Yoshitaka Odagiri, Hideyuki Hasegawa, and Hiroshi Kanai (Grad. School of Eng., Tohoku Univ., Sendai 980-8579, Japan, odagiri@us.ecei.tohoku.ac.jp)

Many studies have been carried out on the measurement of mechanical properties of tissues by applying an ultrasound-induced acoustic radiation force. To assess mechanical properties, strain of an object must be generated. However, one radiation force is not sufficient because it also causes translational motion when the object is much harder than surrounding medium. In this study, two cyclic radiation forces are applied to a muscle phantom from two opposite horizontal directions so that the object is cyclically compressed in the horizontal direction. As a result, the object is vertically expanded due to the incompressibility. The resultant vertical displacement is measured using ultrasound. Two concave ultrasonic transducers for actuation were both driven by sums of two continuous sinusoidal signals at two slightly different frequencies of 1 MHz and (1M +5) Hz. Displacement, which fluctuates at 5 Hz, was measured by the ultrasonic *phased tracking method* proposed by our group. Results indicated that the surface of the phantom was cyclically actuated with an amplitude of a tenth of a few micrometers, which well coincided with that measured with laser vibrometer. In addition, upward and downward displacements at the surface and deeper region were found during the increase phase of radiation forces. Such displacements correspond to the horizontal compression.

3:50

4pBB10. A phantom study on ultrasonic measurement of arterial wall strain combined with tracking of translational motion.
Hideyuki Hasegawa and Hiroshi Kanai (Grad. School of Eng., Tohoku Univ., Aramaki-aza-Aoba 6-6-05, Sendai 980-8579, Japan, hasegawa@us.ecei.tohoku.ac.jp)

Correlation-based techniques are often applied to ultrasonic rf echoes to obtain the arterial wall deformation (strain). In such methods, the displacement estimates are biased due to changes in center frequency of echoes. One of the reasons for the change in center frequency is the interference of echoes from scatterers within the wall. In the phased tracking method previously proposed for strain estimation by our group, the estimated displacement contains both the components due to the translational motion and strain. The translational motion is larger than strain by a factor of 10 and, thus, the error in the estimated displacement due to the change in center frequency mainly depends on translational motion and is often larger than the minute displacement due to strain. To reduce this error, in this study, a method is proposed in which the translational motion is compensated using the displacement of the luminal boundary estimated by the phased tracking method before correlating echoes between the frame before deformation and that at the maximum deformation to estimate the strain distribution within the wall. In basic experiments using phantoms made of silicone rubber, the estimation error was much reduced to 15.6% in comparison with 36.4% obtained by the previous method.

4:05

4pBB11. Wave biomechanics of skeletal muscle.
Oleg Rudenko (Dept. of. Blekinge Inst. of Technol., 371 79 Karlskrona, Sweden) and Armen Sarvazyan (Artann Labs., Inc., West Trenton, NJ 08618)

Physiological functions of skeletal muscle, such as voluntary contraction and force development, are accompanied by dramatic changes of its mechanical and acoustical properties. Experimental data show that during contraction, the muscle's Young's modulus, shear viscosity, and anisotropy parameter are changed by over an order of magnitude. None of the existing models of muscle contraction and muscle biomechanics can adequately explain the phenomena observed. A new mathematical model [O. Rudenko and A. Sarvazyan, Acoust. Phys. (6), (2006)], has been developed relating the shear wave propagation parameters to molecular structure of the muscle and to the kinetics of the mechanochemical crossbridges between the actin and myosin filaments. New analytical solutions describing waves in muscle including nonlinear phenomena are found. A molecular mechanism for the dependence of acoustical characteristics of muscle on its fiber orientation and the contractile state is proposed. It is shown that although the anisotropy connected with the preferential direction along the muscle fibers is characterized by five elastic moduli, only two of these moduli have independent values in the muscle. The potential implications of the proposed model in terms of the acoustical assessment of muscle function are explored.

4pBB12. Phase aberration correction for a linear array transducer using ultrasound radiation force and vibrometry optimization: Simulation study. Matthew W. Urban and James F. Greenleaf (Dept. of Physiol. and Biomed. Eng., Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, urban.matthew@mayo.edu)

Diagnostic ultrasound images suffer from degradation due to tissues with sound speed inhomogeneities causing phase shifts of propagating waves. These phase shifts defocus the ultrasound beam, reducing spatial resolution and image contrast in the resulting image. We describe a phase aberration correction method that uses dynamic ultrasound radiation force to harmonically excite a medium using amplitude-modulated continuous wave ultrasound created by summing two ultrasound frequencies at f_0 $=3.0$ MHz and $f_0 + \Delta f = 3.0005$ MHz. The phase of each element of a linear array transducer is sequentially adjusted to maximize the radiation force and obtain optimal focus of the ultrasound beam. The optimization is performed by monitoring the harmonic amplitude of the scatterer velocity in the desired focal region using Doppler techniques. Simulation results show the ability to regain a 3.0-MHz focused field after applying a phase screen with an rms time delay of 95.4 ns. The radiation force magnitude increased by 22 dB and the resolution of the field was regained. Simulation results show that the focus of the beam can be qualitatively and quantitatively improved with this method. [This study was supported in part by Grants EB002640 and EB002167 from the NIH.]

4pBB13. Application of the optoacoustic technique to visualization of lesions induced by high-intensity focused ultrasound. Tatiana Khokhlova, Ivan Pelivanov, Vladimir Solomatin, Alexander Karabutov (Intl. Laser Ctr., Moscow State Univ., 119992, Moscow, Russia t_khokhlova@ilc.edu.ru), and Oleg Sapozhnikov (Moscow State Univ., 119992, Moscow, Russia)

Today several techniques are being applied to monitoring of high-intensity focused ultrasound (HIFU) therapy, including MRI, conventional ultrasound, and elastography. In this work a new method for noninvasive monitoring of HIFU therapy is proposed: the optoacoustic method. The optoacoustic technique is based on the excitation of wideband ultrasonic pulses through the absorption of pulsed laser radiation in tissue and subsequent expansion of the heated volume. The excited optoacoustic (OA) pulse contains information on the distribution of optical properties within the tissue—light scattering and absorption coefficients. Therefore, if thermal lesions have different optical properties than the untreated tissue, they will be detectable on the OA waveform. The considerable change in light scattering and absorption coefficients after tissue coagulation was measured using techniques previously developed by our group. Heating induced by HIFU also influences the OA signal waveform due to the rise of thermal expansion coefficient of tissue with temperature. This dependence was measured in order to evaluate the feasibility of the OA technique in temperature monitoring. An OA image of HIFU lesion induced by a 1.1 MHz focused transducer in a liver sample was reconstructed using a 64-element wideband array transducer for OA signal detection.

FRIDAY AFTERNOON, 1 DECEMBER 2006

OAHU ROOM, 1:00 TO 3:00 P.M.

Session 4pEAa

Engineering Acoustics: New Electroacoustic Transducers Utilizing Advanced Technolgies and Materials

Juro Ohga, Cochair

Shibaura Inst. of Technology, 3-9-14 Shibaura, Minato-ku, Tokyo 108-8548, Japan

James E. West, Cochair

Johns Hopkins Univ., Dept. of Electrical and Computer Engineering, Barton 105, 3400 N. Charles St., Baltimore, MD 21218-2686

Invited Papers

1:00

4pEAa1. Solid-state photo-microphones or pressure sensors by total reflection. Yasushi Suzuki (Gunma Natl. College 08 Tech., 580, Toriba-cho, Maebashi-shi, Gunma, 371-8530 Japan., suzuki@elc.gunma-ct.ac.jp) and Ken'iti Kido (Tohoku Univ., Yokohama-shi, Kanagawa, 226-0017 Japan)

Solid-state photo-microphones or pressure sensors are proposed. These sensors use a new principle, involving the optical total reflection at the boundary surface between glass and air. The critical angle for total reflection changes by the refractive index of air, which depends on the air density. Sound pressure changes the air density. Therefore, the sound pressure is measurable by detecting the intensity of the reflected light from the total reflection area. The sensitivity of the sensor is investigated theoretically. It is expected that the sensor has sufficient sensitivity for practical use, employing laser light and a curved boundary surface with a large radius of curvature. Some experiments are carried out to verify the theoretical investigations. A He-Ne laser or a laser diode is employed as a light source in the experiments. Experimental results show that the sensor has equivalent sensitivity to that which was theoretically estimated, but that sensitivity is very low. The sensor is useful as a pressure sensor, but it is difficult to realize a microphone for general use at the present. The microphones have no diaphragm and the upper limit in the frequency range is extremely high in principle.

4pEAa2. Micromachined microphones with diffraction-based optical interferometric readout. F. Levent Degertekin (G.W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, levent@gatech.edu), Neal A. Hall (Sandia Natl. Labs, Albuquerque, NM 87185-5800), and Baris Bicen (Georgia Inst. of Technol., Atlanta, GA 30332)

A diffraction-based optical method for integrated interferometric detection of micromachined microphone diaphragm displacement is described. With multichip optoelectronics integration, this approach yields highly sensitive optical microphones in mm-cube volumes. Since the microphone sensitivity does not depend on capacitance, this method changes the paradigm for the backplate and gap structure design. As a result, one can use millimeter size diaphragms to achieve wide frequency response and low thermal mechanical noise levels characteristic of precision measurement microphones. Furthermore, the electrical port of the device, which is freed by optical detection, is used for electrostatic actuation of the microphone diaphragm to tune microphone sensitivity and to generate self-characterization signals. Prototype optical microphone structures have been fabricated using Sandia National Laboratories' silicon based SwIFT-Lite[TM] process. Measurements on these diaphragms show an A-weighted diaphragm displacement noise of 2.4 pm and flat electrostatic response up to 20 kHz. These results indicate the feasibility of realizing measurement microphones with 1.5-mm-diam diaphragms, 15-dBA internal noise, and 40-kHz bandwidth. Application of the detection method in a bio-inspired directional microphone for hearing aids is also discussed. [Work partially supported by NIH Grant 5R01DC005762-03, Sensing and Processing for Hearing Aids.]

4pEAa3. Hardware and software technologies for improvement of hearing characteristics of headphone reproduction. Kiyofumi Inanaga and Yuji Yamada (Audio Codec Development Dept., Technol. Development Group, SONY Corp., Shinagawa Tec., 12-15-3, Tokyo, 108-6201 Japan)

This report specifically describes commercialization technology of a headphone system with out-of-head localization applying dynamic head-related transfer functions (HRTFs) that can localize sound easily over a full 360 deg. A source image by output of conventional headphones is localized inside the listener's head. However, the image can be localized outside the listener's head by wearing headphones over a full 360 deg through accurate simulation of the listener's HRTFs. Developments of headphone systems using signal processing technology for data correction have given rise to the static binaural reproduction system (SBRS). The first part of this speech describes its psychoacoustic characteristics and challenges. A rotating dummy-head that is synchronized with the listener's head movement was produced experimentally to create the dynamic binaural reproduction system (DBRS). Using the DBRS, HRTFs synchronize with the listener's head movement. Psychoacoustic characteristics and advantages of the system are also discussed in this report. Further developments were made to realize the commercialization of the DBRS in areas including piezoelectric gyroscope head-tracking technology, headphone technologies that can reproduce real sound characteristics, and simplification of HRTF signal processing employing a simulator with electronic circuits. Finally, future visions for these technologies will be touched upon.

4pEAa4. Piezoelectret microphones: A new and promising group of transducers. Gerhard M. Sessler and Joachim Hillenbrand (Darmstadt Univ. of Technol., Merckstrasse 25, 64283 Darmstadt, Germany, g.sessler@nt.tu-darmstadt.de)

Piezoelectret microphones, first described a few years ago, are transducers based on the strong longitudinal piezoelectric effect of charged cellular polymers. Such microphones have recently been improved in two respects: Firstly, an expansion process was used to increase the piezoelectric d_{33} coefficients of cellular polypropylene (PP) films in the audio frequency range up to 600 pC/N and, secondly, stacking of several films was applied to increase the microphone sensitivity. Transducers with six films now show open-circuit sensitivities of up to 15 mV/Pa, comparable to that of electret microphones. Other characteristics of piezoelectret microphones are their low equivalent noise level of about 26 dB(A) and the very small total harmonic distortion of less than 0.1% at 140 dB SPL. The piezoelectric activity of the PP films and the microphone sensitivities are stable at room temperature but start to decay above 50 °C. Recently, directional piezoelectret microphones with various directional characteristics have been designed. Major advantages of piezoelectret microphones are their simple design, their low harmonic distortion, and their wide frequency range extending into the ultrasonic region.

4pEAa5. Expansion of frequency range for piezoelectric loudspeakers by new transducer construction. Juro Ohga (Shibaura Inst. of Technol., 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan)

Although simple construction of piezoelectric loudspeakers engenders various merits, expansion of its working frequency range to the very low region is difficult because the mechanically stiff characteristics of conventional piezoelectric ceramic diaphragms prevent their large amplitude operation. This paper proposes two sorts of new piezoelectric loudspeaker construction that are suitable for low-frequency signal radiation. One idea is the use of a tuck-shape diaphragm by a PVDF polymer film bimorph. It has large surface area with a very low resonant frequency. Resonant frequencies and sensitivity frequency characteristics are examined, and control methods of local diaphragm bending are discussed. The other idea is the use of continuous revolution of a piezoelectric ultrasonic motor. It produces a completely controlled large output force because its output mechanical impedance is much greater than that of any conventional transducer or motor. An ultrasonic motor, whose stator is connected to a direct-radiator loudspeaker cone by a rod and whose rotor is burdened by a heavy metal ring, rotates with a constant velocity. Modulation of the velocity by using an audio signal imparts a driving force to the diaphragm because the heavy ring tends to keep a constant velocity. Experimental models suggest that this construction is useful.

4pEAa6. Modal array signal processing using circular microphone arrays applied to acoustic source detection and localization problems. Heinz Teutsch (Avaya Labs, 233 Mt. Airy Rd., Basking Ridge, NJ 07920, teutsch@avaya.com) and Walter Kellermann (Univ. of Erlangen-Nuremberg, Erlangen, Germany)

Many applications of acoustic signal processing rely on estimates of several parameters present in the observed acoustic scene such as the number and location of acoustic sources. These parameters have been traditionally estimated by means of classical array signal processing (CASP) algorithms using microphone arrays. Algorithms for parameter estimation solely based on the paradigm of CASP often suffer from the narrowband assumption underlying the signal model. This restriction limits their usability when wideband signals, such as speech, are present in the wave field under observation. We investigate the parameter estimation problem by applying the notion of wave field decomposition using baffled circular microphone arrays. The obtained wave field representation is used as the basis for "modal array signal processing algorithms." It is shown that by applying the notion of modal array signal processing, novel algorithms can be derived that have the potential to unambiguously detect and localize multiple simultaneously active wideband sources in the array's full field-of-view. Performance evaluations by means of simulations, measurements, and real-time case studies are presented.

FRIDAY AFTERNOON, 1 DECEMBER 2006　　　　　　　　　　　　　　OAHU ROOM, 3:15 TO 6:00 P.M.

Session 4pEAb

Engineering Acoustics: Special Topics in Engineering Acoustics

Timothy W. Leishman, Cochair
Brigham Young Univ., Dept. of Physics and Astronomy, N247 ESC, Provo, UT 84602

Kiyofumi Inanaga, Cochair
Sony Corp., Shinagawa Tec. 12-15-3, Tokyo 108-6201, Japan

Contributed Papers

3:15

4pEAb1. Enhanced voided piezoelectric polymer for underwater acoustic sensors. Juan Arvelo (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099), Ilene Busch-Vishniac, and James West (Johns Hopkins Univ., Baltimore, MD 21218)

A charged voided polymer has been shown to exhibit large piezoelectricity. This material consists of injected air bubbles into polypropylene. This sheet of voided material is then biaxially stretched to elongate the voids. After stretching this material, a strong electric field is applied to cause dielectric breakdown of the gas in the voids, creating electric charges that are trapped in the polymer frame. Since the sides of the voids have opposite charges, they form macroscopic dipoles. When an external force is applied to this material, the voids become narrower, causing stronger dipole strength. A simple model of this voided material was implemented to derive formulas to estimate its piezoelectric constant, electro-mechanical coupling factor, resonance frequency, and sensor sensitivity based on electrical and mechanical properties of the polymer and gas in the voids. These formulas and a survey of available polymers and gases yielded promising combinations that result in more sensitive voided materials that satisfy selected criteria. These criteria include high sensitivity and maximum service temperature, low dissipation factor, and high dynamic compressibility, but low hydrostatic compressibility. This talk will describe the model, derive the formulas, uncover measured properties of candidate polymers and gases, and show calculated sensitivity of selected polymer/gas combinations.

3:30

4pEAb2. Basic study on one-dimensional transducer array using hydrothermally synthesized lead zirconium titanete poly-crystalline film. Akito Endo, Tomohito Hasegawa, Norimichi Kawashima, Shinichi Takeuchi (1614, Kurogane-cho, Aoba-ku, Yokohama, Kanagawa, 225-8502, Japan), Mutsuo Ishikawa, and Minoru Kurosawa (Midori-ku, Yokohama, Kanagawa 226-8502, Japan)

Recently, high-frequency miniature medical ultrasound probes with high resolution were actively developed. However, it is difficult to fabricate such tiny ultrasound probes using piezoelectric ceramic vibrator with thickness less than 100 μm. We deposited a PZT poly-crystalline film on a titanium substrate using the hydrothermal method and developed transducers using the PZT poly-crystalline film for ultrasound probes. In this study, we applied it to a miniature medical one-dimensional (1-D)-array-type ultrasound probe with resonance frequency of 10 MHz. After sputtering of pure titanium on the surface of a hydroxyapatite substrate, the titanium film was etched using the photolithography method to form a 1-D titanium film electrode array with 75 μm element pitch, 40 μm element width, and 4 mm element length to scan an ultrasound beam electronically by sector scan mode using phased-array technique. Thereby we fabricated a miniature 1-D-array-type ultrasound probe. A transmitted ultrasound pulse from 10 MHz commercial ultrasound probe was received by this fabricated 1-D-array type ultrasound probe with hydrothermally synthesized PZT poly-crystalline film vibrators.

3:45

4pEAb3. Analysis of a barrel-stave flextensional transducer using MAVART (model to analyze the vibrations and acoustic radiation of transducers) and ATILA (analysis of transducers by integration of LAplace equations) finite-element codes. Richard A. G. Fleming, Mark Kwiecinski, and Dennis F. Jones (Defence R&D Canada—Atlantic, P.O. Box 1012, Dartmouth, NS, Canada B2Y 3Z7, dennis.jones@drdc-rddc.gc.da)

A small barrel-stave flextensional transducer, designed and tested at Defence Research and Development Canada—Atlantic, is a candidate sound source for underwater coastal surveillance and acoustic communications applications. This high-power transducer has an outside diameter, length, and mass of 5.7 cm, 12.7 cm, and 1.1 kg, respectively. The measured fundamental flexural resonance frequency was 1.8 kHz with a transmitting voltage response of 118 dB/1μPa-m/V and an omnidirectional radiation pattern. Two finite-element models were developed for this transducer using the finite-element codes MAVART (Model to Analyze the Vibrations and Acoustic Radiation of Transducers) and ATILA (Analysis of Transducers by Integration of LAplace equations). Comparisons are made between the calibration measurements and the model predictions. [Work supported in part by Sensor Technology Limited.]

4:00

4pEAb4. Thermal behavior of high-power active devices with the ATILA (analysis of transducers by integration of LAplace equations) finite-element code. Jean-Claude Debus (Institut Superieur de l'Electronique et du Numerique, 41 Bv Vauban, 59046 Lille, Cedex France), John BlottmanIII, and Stephen Butler (Naval Undersea Warfare Ctr. Div. Newport, RI 02841)

Many active devices using piezoelectric ceramics are driven with very high power densities and long pulse lengths. Due to mechanical and dielectric losses in the materials, this produces heat, causing a temperature rise in the devices, which may lead to their mechanical failure. The thermal issues have been shown to be the limiting device design criteria over electric field and mechanical stress limits, yet the effect of the temperature on performance is generally not considered in the numerical models used during the design stage. A coupled electro-mechanical thermal analysis is implemented in the ATILA code. For a steady-state or transient solution, a thermal behavior is weakly coupled to the electromechanical response. The method may take advantage of the order-of-magnitude-greater time constant for thermal effects compared to mechanical behavior. A two-step analysis is performed whereby the electromechanical behavior is first computed, and the resulting dissipated power is then applied as a heat generator to determine the resulting temperature of the device. A high-drive, 31-mode, free flooded ring transducer and a sonar projector serve as validation of the numerical model. The approach addresses both the transient thermal response and the steady temperature profile that results from the high-power, high-duty-cycle drive.

4:15

4pEAb5. Development of multichannel optical sensor and visualization of vibration distribution. Jun Hasegawa and Kenji Kobayashi (Faculty of Eng., Takushoku Univ., 815-1 Tatemachi, Hachioji-shi, Tokyo 193-0985 Japan, jhase@es.takushoku-u.ac.jp)

A multi-channel optical sensor system was developed to measure vibrations simultaneously with high spatial resolution. As sensor elements, optical displacement sensor units were developed not to disturb the natural vibration. Each sensor unit, which consists of the optical fiber bundle and focusing lens, can detect the displacement of the object as the variation of the reflected light power. The sensor unit has a displacement resolution of 10 nm, a dynamic range of more than 90 dB, and a frequency band width of up to 80 kHz. Up to 64 sensor units can be arrayed as one sensor head, which realizes the simultaneous measurement of vibration distribution with the high spatial resolution of 4 mm. A calibrating function under the measurement circumstances was developed. Under calibration mode, the sensor array head is moved by a linear actuator, while the vibration of the object is stopped. Thus the calibrated data of each sensor unit can be

obtained for the displacement magnitude. Measured vibration distributions can be monitored as the three-dimensional animations. With the system developed, several actuators for vibratory micro-injection were measured, and the system could reveal their detailed vibration distributions and could detect the existence of a failure portion of some actuator.

4:30

4pEAb6. Prediction of howling for a sound system in an acoustical environment with both reverberant and direct sounds. Hideki Akiyama and Juro Ohga (Shibaura Inst. of Technol., 3-7-5 Toyosu, Koto-ku, Tokyo 135-8548, Japan, m106003@shibaura-it.ac.jp)

Prediction of howling is a key technology for a howling suppression design for a sound system with a loudspeaker and microphone. A howling occurrence prediction method for a sound system in a reverberant room has already been presented [J. Ohga, J. Sakaguchi, "Prediction of howling of a sound system in a reverberant room," W. C. Sabine Centennial Symposium (ASA, New York, 1994), 2aAAd4]. It is apparently useful for ordinary public address systems whose distances of loudspeakers from microphones are large. However, this result was not perfect because the direct sound component is not negligible in hands-free telephones or teleconference systems whose loudspeakers and microphones are set close to each other. This report gives a quantitative howling occurrence prediction method for a sound system in an acoustical environment with both reverberant and direct sounds. The following design parameters are obtained: (1) the increase of howling occurrence level from the power average value, (2) the level occurrence probability, and (3) the critical level chart given by an equation as a function of direct and reverberant sounds ratio. Prediction results for particular examples are compared with calculations of sound-field transfer functions. Results confirmed that it is practical.

4:45

4pEAb7. Effect of background noise on dialogue in telephony. Koichi Amamoto and Juro Ohga (Sibaura Inst. of Technol., 3-7-5 Toyosu, Koto-ku, Tokyo, 135-8548, Japan, m106006@shibaura-it.ac.jp)

Recent developments of mobile telephones include new sorts of impairments against speech. Conventional evaluation method for impairments by a talker and a few listeners cannot apply to these new ones, because they are brought by long signal delay. The effect of it cannot discriminate by "one-sided" test. This research relates to a speech quality evaluation by conversation between two persons. Variation of conversation stream is observed by addition of pink noise of various levels to a dialogue by microphones and earphones. Length of sentences and frequency of repeats are quantified and their meanings are discussed

5:00

4pEAb8. Best practices for auditory alarm design in space applications. Durand Begault and Martine Godfroy (Human Systems Integration Div., NASA Ames Res. Ctr., Moffett Field, CA 94035)

This presentation reviews current knowledge in the design of auditory caution and warning signals, and sets criteria for development of "best practices" for designing new signals for NASA's Crew Exploration Vehicle (CEV) and other future spacecraft, as well as for extra-vehicular operations. A design approach is presented that is based upon cross-disciplinary examination of psychoacoustic research, human factors experience, aerospace practices, and acoustical engineering requirements. Existing alarms currently in use with the NASA Space Shuttle flight deck are analyzed and then alternative designs are proposed that are compliant with ISO 7731 ("Danger signals for work places Auditory Danger Signals") and that correspond to suggested methods in the literature to insure discrimination and audibility. Parallel analyses are shown for a sampling of medical equipment used in surgical, periop, and ICU contexts. Future development of auditory sonification techniques into the design of alarms will allow auditory signals to be extremely subtle, yet extremely useful in subtly indicating trends or root causes of failures. [Work funded by NASA's Space Human Factors Engineering Project.]

5:15

4pEAb9. Acoustic signal analysis for forensic applications. Durand Begault and Christopher Peltier (Audio Forensic Ctr., Charles M. Salter Assoc., 130 Sutter St., Ste. 500, San Francisco, CA 94104, durand.begault@cmsalter.com)

Acoustical analysis of audio signals is important in many legal contexts for determining the authenticity, originality, and continuity of recorded media; determining the circumstances of events in question that may have been recorded; for determining the audibility of signals; and for identification or elimination of talkers as a match to an unknown exemplar. Recorded media are analyzed in forensic applications using both familiar techniques (waveform and spectral analyses) and more novel methods (e.g., ferro fluid development of media; specialized tape heads with non-standard reproduction characteristics; crystal microscopy; detection and matching to power grid frequencies). Audibility analyses frequently require careful reconstructive field measurements and criteria in excess of normally accepted standards. Voice identification-elimination protocols must account for examiner bias and exemplar quality and can be described using a receiver operator curve (ROC) model. This presentation gives an overview of these techniques and their comparative advantages for specific forensic applications.

5:30

4pEAb10. Without low-pass filter for the 1-bit digital amplifier. Kiyoshi Masuda (Coroprate Res. and Development Group, SHARP Corp., 2613-1 Ichinomoto-cho, Tenri-shi, Nara, Japan) and Yoshio Yamasaki (Waseda Univerity, Okubo, Shinzyuku-ku, Tokyo, Japan)

SHARP collaborated with Waseda University from 1990 for 1-bit digital technology. SHARP had started to receive an order for the 1-bit digital amplifier "SM-SX100" on 20 August 1999. Until today, we have introduced the 1-bit digital amplifier for audio, flat panel TV (LCD TV), and PC. These 1-bit amplifiers provided low-pass filter for the final stage,

which is provided after 1-bit digital switching. We have to achieve more good sound and reduce deterioration of this low-pass filter. We have introduced new 1-bit digital amplifier without this low-pass filter beginning this April. This means we controlled the 1-bit digital signal to directly operate the speaker. We have proved a better effect for sound to compare the new 1-bit digital amplifier with the PWM switching amplifier, the A-class amplifier and the 1-bit digital amplifier with low-pass filter. If we do not measure any improvement for this new 1-bit digital amplifier, it has large radiation noise. We had achieved a reduction to the limit level of FCC, Denanhou, etc.

5:45

4pEAb11. Force-frequency effect of thickness mode langasite resonators. Haifeng Zhang (W317.4 Nebraska Hall, Univ. of Nebraska, Lincoln, NE 68588-0526, hfzhang@bigred.unl.edu), Joseph A. Turner, Jiashi Yang (Univ. of Nebraska, Lincoln, NE 68588-0526), and John A. Kosinski (U.S. Army CECOM, Fort Monmouth, NJ 07703-5211)

Langasite resonators are of recent interest for a variety of applications because of their good temperature behavior, good piezoelectric coupling, low acoustic loss, and high Q factor. The force-frequency effect describes the shift in resonant frequency a resonator experiences due to the application of a mechanical load. A clear understanding of this effect is essential for many design applications such as pressure sensors. In this presentation, the frequency shift is analyzed theoretically and numerically for thin, circular langasite plates subjected to a diametrical force. The results are compared with experimental measurements of the same system for a variety of langasite resonators with various material orientations. In addition, the sensitivity of force-frequency effect is analyzed with respect to the nonlinear material constants. A comparison between the force-frequency effect of langasite and quartz resonators is also made. Finally, the application of such measurements for determining third-order elastic constants is discussed. [Work supported by ARO.]

FRIDAY AFTERNOON, 1 DECEMBER 2006   IAO NEEDLE/AKAKA FALLS ROOM, 1:20 TO 4:25 P.M.

Session 4pMU

Musical Acoustics and Psychological and Physiological Acoustics: Acoustic Correlates of Timbre in Musical Instruments

James W. Beauchamp, Cochair

Univ. of Illinois Urbana-Champaign, School of Music, Dept. of Electrical and Computer Engineering, 1002 Eliot Dr., Urbana, IL 61801

Mashashi Yamada, Cochair

Kanazawa Inst. of Technology, Dept. of Media Informatics, 3-1 Yatsukaho, Hakusan, Ishikawa 924-0838, Japan

Invited Papers

1:20

4pMU1. A meta-analysis of acoustic correlates of timbre dimensions. Stephen McAdams, Bruno Giordano (CIRMMT, Schulich School of Music, McGill Univ., 555 Sherbrooke St. West, Montreal, QC, Canada H3A 1E3), Patrick Susini, Geoffroy Peeters (STMS-IRCAM-CNRS, F-75004 Paris, France), and Vincent Rioux (Maison des Arts Urbains, F-75020 Paris, France)

A meta-analysis of ten published timbre spaces was conducted using multidimensional scaling analyses (CLASCAL) of dissimilarity ratings on recorded, resynthesized, or synthesized musical instrument tones. A set of signal descriptors derived from the tones was drawn from a large set developed at IRCAM, including parameters derived from the long-term amplitude spectrum (slope, centroid, spread, deviation, skewness, kurtosis), from the waveform and amplitude envelope (attack time, fluctuation, roughness), and from variations in the short-term amplitude spectrum (flux). Relations among all descriptors across the 128 sounds were used to determine families of related descriptors and to reduce the number of descriptors tested as predictors. Subsequently multiple correlations between descriptors and the positions of timbres along perceptual dimensions determined by the CLASCAL analyses were

computed. The aim was (1) to select the subset of acoustic descriptors (or their linear combinations) that provided the most generalizable prediction of timbral relations and (2) to provide a signal-based model of timbral description for musical instrument tones. Four primary classes of descriptors emerge: spectral centroid, spectral spread, spectral deviation, and temporal envelope (effective duration/attack time). [Work supported by CRC, CFI, NSERC, CUIDADO European Project.]

1:40

4pMU2. Perceptual acoustics of consonance and disssonance in multitimbral triads. Roger Kendall (Music Cognition and Acoust. Lab., Program in Systematic Musicology, UCLA, 405 Hilgard Ave., Los Angeles, CA 90095, kendall@ucla.edu) and Pantelis Vassilakis (DePaul Univ., Chicago, IL 60614)

Most studies of consonance and dissonance assume a singular spectrum for the constituent intervals of a dyad. Recently, the principal author conducted experiments evaluating triads consisting of digitally mixed combinations drawn from the MUMS single-note natural-instrument recordings. Results indicated that the main effect of ratings for consonance and dissonance correlated well with studies using artificial signals. However, interaction effects suggested perceptual differences related to the timbral differences across combinations. The present experiment evaluates perceptual and acoustical variables of the ten possible triadic combinations created with C4 as the lower and the ten with C5 as the upper notes. UCLA wind ensemble performers on oboe, flute, and clarinet, combinations designed to span timbral space, were digitally recorded. Analyses include perceptual ratings of consonance and dissonance, similarity, as well as acoustical analysis of roughness using a recently developed model. Since natural performances of any type vary in fundamental frequency, additional experiments will employ emulated oboe, flute, and clarinet (using the Kontakt Silver synthesizer in Sibelius 4) as well as purely synthetic stimuli, in order to ascertain the relationship of time-variant spectral properties to consonance, dissonance, and perceived similarity.

2:00

4pMU3. Multidimensional scaling analysis of centroid- and attack/decay-normalized musical instrument sounds. James W. Beauchamp (School of Music and Dept. of Elect. & Comput. Eng., Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801, jwbeauch@uiuc.edu), Andrew B. Horner (Hong Kong Univ. of Sci. & Technol., Kowloon, Hong Kong), Hans-Friedrich Koehn, and Mert Bay (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Ten sustained musical instrument tones (bassoon, cello, clarinet, flute, horn, oboe, recorder, alto saxophone, trumpet, and violin) were spectrally analyzed and then equalized for duration, attack and decay time, fundamental frequency, number of harmonics, average spectral centroid, and presentation loudness. The tones were resynthesized both with time-varying harmonic amplitudes and frequencies (dynamic case) and fixed amplitudes and frequencies (static case). Tone triads were presented to ten musically experienced listeners whose tasks were to specify the most dissimilar and most similar pairs in each triad. Based on the resulting dissimilarity matrix, multidimensional scaling (MDS) was used to position the instruments in two- and three-dimensional metric spaces. Two measures of instrument amplitude spectra were found to correlate strongly with MDS dimensions. For both the static- and dynamic-case 2-D solutions, the ratio of even-to-odd rms amplitudes correlated strongly with one of the dimensions. For the dynamic case, spectral centroid variation correlated strongly with the second dimension. Also, 2-D solution instrument groupings agreed well with groupings based on coefficients of the first two components of a principle components analysis representing 90% of the instruments' spectral variance. [This work was supported by the Hong Kong Research Grants Council's CERG Project 613505.]

2:20

4pMU4. Sound synthesis based on a new micro timbre notation. Naotoshi Osaka (School of Eng., Tokyo Denki Univ., 2-2, Kanda-Nishikicho, Chiyoda-ku, Tokyo, 101-8457, Japan, osaka@im.dendai.ac.jp), Takayuki Baba, Nobuhiko Kitawaki, and Takeshi Yamada (Univ. of Tsukuba, Japan)

Timbre has become a major musical factor in contemporary and computer music. However, sufficient timbre theory has not yet been established. The author is challenging to create new timbre theory for music composition. The first step of its construction is to make the timbre descriptive. A micro timbre is defined, which is a perceptual impression of a sound with approximately 50 to 100-ms duration, and describe sound as a micro timbre sequence. This can be used as a new notation system in place of common music notation. In dictation, micro timbre sequence and correspondent duration sequence are perceptually recorded. When synthesizing from this notation, sounds corresponding to the notation systems are either physically synthesized or searched for in a large sound database to generate sound data for a given duration. Two sequential sound data instances are first represented in sinusoidal representations and then are concatenated using a morphing technique. Sounds generated by a stream of water and similar sounds are described using the method as examples. Then scripts describing electronic sounds are introduced and explained. The ability to record, transmit to others, and resynthesize timbre is one of the useful functions of the theory.

2:40

4pMU5. Timbre representation for automatic classification of musical instruments. Bozena Kostek (Gdansk Univ. of Technol., Narutowicza 11/12, PL-80-952 Gdansk, Poland)

Human communication includes the capability of recognition. This is particularly true of auditory communication. Music information retrieval (MIR) turns out to be particularly challenging, since many problems remain still unsolved. Topics that should be included within the scope of MIR are automatic classification of musical instruments/phrases/styles, music representation and indexing, estimating musical similarity using both perceptual and musicological criteria, recognizing music using audio and/or semantic description, language modeling for music, auditory scene analysis, and others. Many features of music content description are based on perceptual phenomena and cognition. However, it can easily be observed that most of the low-level descriptors used, for example, in musical instrument classification are more data- than human-oriented. This is because the idea behind these features is to have data

defined and linked in such a way as to be able to use it for more effective automatic discovery, integration, and reuse in various applications. The ambitious task is, however, to provide seamless meaning to low- and high-level descriptors such as timbre descriptors and linking them together. In such a way data can be processed and shared by both systems and people. This paper presents a study related to timbre representation of musical instrument sounds.

3:00−3:15   Break

3:15

4pMU6. An attempt to construct a quantitative scale of musical brightness for short melodies implementing timbral brightness. Masashi Yamada (Kanazawa Inst. of Technol., 3-1 Yatsukaho, Hakusan, Ishikawa 924-0838, Japan, m-yamada@neptune.kanazawa-it.ac.jp)

It is known that a major tune is brighter than a minor one, and that music played in a faster tempo and a higher register is brighter than a slower and lower one. However, it has not been clarified how these factors quantitatively determine the musical brightness. On the other hand, it has been clarified that the timbral brightness of a tone corresponds well to the spectral centroid. In the present study, major and minor scales and two short melodies were played with pure tones, and listeners evaluated their musical brightness. For each performance, the spectral centroid was calculated for the overall-term spectrum during the performance on the transformed frequency scale of the ERB rate. The results showed that the musical brightness of the ascending scale increases proportionally as the spectral centroid shown in the ERB rate increases. Using this, a quantitative scale of musical brightness, BM, was constructed. The results also showed that the difference in the musical brightness between major and minor scales corresponded to the transposition of approximately 5 ERB rate, and doubling the speed corresponded to the upper shift of the centroid in approximately 2.5 ERB rate.

3:35

4pMU7. Subjective congruency between a sound effect and a switching pattern of a visual image. Shinichiro Iwamiya, Motonori Arita, and Sun Su (Dept. of Acoust. Design, Kyushu Univ., 4-9-1, Shiobaru, Minami-ku, Fukuoka 185-8540, Japan)

The relationship between the transformation of a visual image and the pitch pattern of sound can create formal congruency between sounds and moving pictures. This effect by switching patterns of enlarging and reducing images that were combined with ascending and descending pitch scales was examined. Rating experiments showed two congruent patterns of the combination of switching and scale patterns; one was a combination of an ascending pitch scale and an enlarging image pattern, and the other a combination of a descending pitch scale and a reducing image pattern. These forms of matching might be based on a Doppler illusion. An additional pair of congruent patterns for combinations of switching and scale patterns was also found: one was a combination of an ascending pitch scale and a sliding movement from left to right, and the other a combination of a descending pitch scale and a sliding movement from right to left. These forms of matching might be based on the correspondence of a progressive sensation. Further, the formal congruency between a pitch pattern and the formal transportation can contribute to integrating auditory and visual information and to making audio-visual products more impressive.

Contributed Papers

3:55

4pMU8. SRA: An online tool for spectral and roughness analysis of sound signals. Pantelis Vassilakis (School of Music, ITD, Libraries, DePaul Univ., 2350 N. Kenmore Ave., JTR 207, Chicago, IL 60614)

SRA performs spectral and roughness analysis on user-submitted 250- to 1000-ms-long portions of sound files (.wav/.aif formats). Spectral analysis incorporates an improved STFT algorithm [K. Fitz and L. Haken, J. Aud. Eng. Soc. **50**(11), 879–893 (2002)] and automates spectral peak-picking using the Loris open source C++ class library [Fitz and Haken (CERL Sound Group)]. Users can manipulate three spectral analysis/peak-picking parameters: analysis bandwidth, spectral-amplitude normalization, and spectral-amplitude threshold. Instructions describe the parameters in detail and suggest settings appropriate to the submitted files and questions of interest. The spectral values obtained from the analysis enter a roughness estimation model [P. N. Vassilakis, Sele. Rep. in Ethnomusicol. **12**, 119–144 (2005)], outputting roughness values for each individual sine-pair in the file's spectrum and for the entire file. The roughness model quantifies the dependence of roughness on a sine-pair's (a) intensity (combined amplitude of the sines), (b) amplitude fluctuation degree (amplitude difference of the sines), (c) amplitude fluctuation rate (frequency difference of the sines), and (d) register (lower sine frequency). Presentation of the roughness estimation model and the online tool will be followed by a discussion of research studies employing it and an outline of future possible applications. [Work supported by DePaul University and Eastern Washington University. Programmed by K. Fitz.]

4:10

4pMU9. Further spectral correlations of timbral adjectives used by musicians. Alastair C. Disley, David M. Howard, and Andrew D. Hunt (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK)

As part of a project to develop a synthesis interface which nontechnical musicians should find intuitive, the adjectives musicians use to describe timbre have been studied in a large-scale listening test covering the timbre space of Western orchestral instruments. These were refined in previous work by the authors (A. C. Disley *et al.* "Spectral correlations of timbral adjectives used by musicians," J. Acoust. Soc. Am. **119**, 3333, (2006)] to a set of ten words which had good common understanding and discrimination between the samples (bright, clear, dull, gentle, harsh, nasal, percussive, ringing, thin, and warm). To help explore potential relationships between these adjectives and spectral features, 20 listeners participated in a further listening experiment, comparing samples in pairs to produce dissimilarity data. Multidimensional scaling produced dimensions which were compared with a large number of spectral and time-domain analyses of the stimuli, suggesting a number of significantly correlated spectral cues with some of the adjectives. These results are compared with previous studies by the authors and others, showing both similarities and differences, suggesting that collective consideration of timbral adjectives is more likely to result in simultaneously applicable theories of acoustic correlates than individual consideration of words.

Session 4pNSa

Noise and Architectural Acoustics: Soundscapes and Cultural Perception II

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Technical Acoustics, Secr TA 7, Einsteinufer 25, 10587 Berlin, Germany

Bennett M. Brooks, Cochair
Brooks Acoustics Corp., 27 Hartford Turnpike, Vernon, CT 06066

Contributed Papers

1:00

4pNSa1. Mapping soundscapes in urban quiet areas. Gaetano Licitra (ARPAT, Tuscany Regional Agency for Environ. Protection, Via N. Porpora, 22-50144, Firenze, Italy), Gianluca Memoli (Memolix, Environ. Consultants, 56127 Pisa, Italy), Mauro Cerchiai, and Luca Nencini (ARPAT, 56127 Pisa, Italy)

Innovative action plans in noise-polluted environments require the description of the existing soundscape in terms of suitable indicators. The role of these indicators, giving the "fingerprint" of a fixed soundscape, would be not only to measure the improvement in the sound quality after the action taken, but also to guide the designer in the process, providing a reference benchmark. One of the open questions on new indicators is the way they relate to existing ones and to people's perception. The present work will describe a "Sonic Garden" in Florence, using both the "slope" indicator (constructed from the LA_{eq} time history and related in previous studies to people's perception) and classical psychoacoustical parameters (level, spectral structure, and perceived characteristics such as loudness, sharpness, fluctuation, and roughness). The latter parameters will be acquired using a binaural technique.

1:15

4pNSa2. A questionnaire survey of the attitude of Japanese and foreign residents in Japan to sound masking devices for toilets. Miwako Ueda and Shin-ichiro Iwamiya (Grad. School of Design, Kyushu Univ., Iwamiya Lab. 4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540 Japan, amaria@white.livedoor.com)

Unique sound masking devices for toilets can be used in women's restrooms in Japan. Such devices function to produce the sound of flushing water without actual flushing. To mask the sound of bodily functions, women tended to flush the toilet continuously while using it, thereby wasting a large amount of water. In the circumstances, sound masking devices have been introduced to public toilets. We have recently conducted a questionnaire survey to clarify the attitude of people toward such sound masking devices for toilets. The results of the survey showed that many Japanese women know such devices and often use them, that foreign women currently living in Japan also know that such devices exist, and that some Japanese men have heard of such devices but never used them. Many Japanese women are quite embarrassed at the thought that someone else can hear them while they are on the toilet. Many noted the necessity of such devices and required a wide range of setting for toilets in public spaces. However, they are not satisfied with the sound quality of the playback toilet flush sounds of currently available devices. The above results suggest that the sound quality of such devices should be improved.

1:30–2:00
Panel Discussion

Session 4pNSb

Noise, Physical Acoustics, and Structural Acoustics and Vibration: Acoustics of Sports

Joseph Pope, Cochair
Pope Engineering Company, P.O. Box 590236, Newton, MA 02459-0002

Kenji Kurakata, Cochair
AIST, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

Chair's Introduction—2:10

Invited Papers

2:15

4pNSb1. A review of the vibration and sounds of the crack of the bat and player auditory clues. Robert Collier (Thayer School of Eng., 8000 Cummings Hall, Hanover, NH 03755)

The purpose of this paper is to review the state-of-the-art in the acoustics of baseball. As is well known, the crack of the bat is an important phenomenon of solid wood bats and metal bats. Each has a very different sound signature. At the 148th meeting of the ASA in 2004, the author and coauthors Ken Kaliski and James Sherwood presented the results of laboratory and field tests, which showed

that the spectral characteristics of radiated sound are dependent on the ball-bat impact location and resultant bat vibration of both solid wood and tubular metal bats. These results will be reviewed together with those of other investigators in the context of player auditory clues and the player's response in game situations.

2:35

4pNSb2. Measurements of the impact sound of golf clubs and risk of hearing impairment. Kenji Kurakata (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki, 305-8566 Japan, kurakata-k@aist.go.jp)

The ball-impact sounds of golf clubs with metal heads and of a club with a wood head were measured to investigate their different acoustic properties. Hitting was executed using either a swing machine or a human player. Results of these analyses showed that the metal-head clubs generated sounds around 100 dB ($L_{pA,Fmax}$). This level was 5–15 dB higher than that of the wood-head club. The sounds of the metal-head clubs had greater power in the high-frequency region of 4 kHz and above compared to the wood-head club, which particularly increased the overall sound levels. These results suggest that it would be desirable to develop a metal head with pleasant sound qualities, keeping the sound level lower to minimize hearing damage. Some of these measurement data were published in Japanese in a previous paper [K. Kurakata, J. INCE/J **26**, 60–63 (2002)].

2:55

4pNSb3. New underwater sound system for synchronized swimming: The 9th International Swimming Federation Championships. Takayuki Watanabe, Shinji Kishinaga (YAMAHA Ctr. for Adv. Sound Technologies, 203 Matsunokijima, Iwata, Shizuoka, 438-0192 Japan), Tokuzo Fukamachi (YAMAHA Motor Marine Operations, Arai, Hamana, 438-8501 Japan), and Osamu Maeda (YAMAHA Motor Adv. Technol. Res. Div., Iwata, 438-8501 Japan)

There have been concerns about the differences between underwater sound fields in a temporary fiberglass-reinforced plastic (FRP) pool and in a conventional reinforced concrete (RC) pool. A temporary FRP pool was to be used for competitions at the World Swimming Championships in Fukuoka. We considered three items as key factors for a swimming pool used for synchronized swimming: (1) the sound source itself (output level, fluctuations in frequency characteristics); (2) the effect of materials used in pool construction upon sound source installation conditions; and (3) the effect of the mth mode low-frequency cutoff in "shallow water." To improve basic problems related to the first factor, we developed a new actuator-driven underwater sound system (YALAS), which can eliminate the effect of installation conditions for underwater speakers in the FRP pool. This new underwater system has now seen practical use in competitions. The report summarizes this new underwater sound system and compares the system with conventional systems in terms of its acoustic characteristics. The system can offer music with sufficient audibility in water. We gained a good reputation with competitors because the system showed superior performance to conventional systems in sound volume and quality, and in uniformity of sound distribution.

3:15

4pNSb4. Acoustics of the Great Ball Court at Chichen Itza, Mexico. David Lubman (14301 Middletown Ln., Westminster, CA 92683)

The ball game has played a central role in Mayan religion and culture for 5000 years. Thousands of ball courts have been discovered. The Great Ball Court (GBC) at Chichen Itza is a late development and is architecturally unique. Two remarkable acoustical features were noticed during excavation in the 1920s, but never explained or interpreted. A whispering gallery permits voice communication between temples located about 460 feet (140 m) apart. A profound flutter echo is heard between the two massive parallel walls of the playing field, about 270 ft (82 m) long, 28 ft (8.5 m) high, and 119 ft (36 m) apart. Until recently, most archaeologists dismissed acoustical features at Mayan sites as unintended artifacts. That is now changing. Stimulated by archaeological acoustic studies and reports since 1999, eminent Mayanists Stephen Houston and Karl Taube have reinterpreted certain Mayan glyphs as vibrant sounds and ballcourt echoes, and have famously called for a new archaeology of the senses, especially hearing, sight, and smell [Cambridge Archaeol. J. **10** (2) 261–294 (2000)]. By interpreting architectural, psychoacoustic, and cognitive features of the GBC in the context of ancient Mayan culture, this paper speculates that acoustical effects at the GBC may be original design features.

Contributed Papers

3:35

4pNSb5. Sleep disturbance caused by shooting sounds. Joos Vos (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands, joos.vos@tno.nl)

In the present study relations between the sound level of shooting sounds and the probability of behaviorally confirmed noise-induced awakening reactions were determined. The sounds were presented by means of loudspeakers in the bedrooms of 30 volunteers. The shooting sounds had been produced by a small and a medium-large firearm, and the stimuli consisted of individual bangs or volleys of ten isolated or partly overlapping impulses. Aircraft sound was included as a reference source. The sounds were presented during a 6-h period that started 75 min after the beginning of the sleeping period. The time period between the various stimuli varied between 12 and 18 min, with a mean of 15 min. To cope with at least a relevant portion of habituation effects, each subject participated in 18 nights to be completed within 4 weeks. Preliminary results are presented both for the awakening reactions described above, and for vari-

ous other dependent variables collected with the help of an actimeter or determined by means of subjective rating scales. [Work supported by the Dutch Ministry of Defense.]

3:50

4pNSb6. Sound inside a gymnasium. Sergio Beristain (ESIME, IPN, IMA., P.O. Box 12-1022, Narvarte, 03001, Mexico City, Mexico)

A new gymnasium for a sports club was designed taking acoustic confort into consideration, in order to accomodate sports practice, sports events with the public, or musical and drama presentations, taking advantage of its large capacity for the public and performers. The floor plan included room enough for a basketball court with public space on one side, where grades for about 200 people will be permanently installed. Walls were treated in a way that is useful for the sports practice (hard surfaces), with hidden absorption material to reduce the usual reverberant

field inside the court, and to allow for sound events with only the addition of a mat (to protect the floor woodwork) and extra grades and the sound reinforcement system.

4:05

4pNSb7. Occupational and recreational noise exposures at stock car racing circuits. Chucri A. Kardous, Thais Morata, and Luann E. Van Campen (Natl. Inst. for Occupational Safety Health, 4676 Columbia Pkwy., Cincinnati, OH 45226, ckardous@cdc.gov)

Noise in stock car racing is accepted as a normal occurrence but the exposure levels associated with the sport have not been adequately characterized. Researchers from the National Institute for Occupational Safety and Health (NIOSH) conducted an exploratory assessment of noise exposures to drivers, pit crew, team staff, and spectators at three stock car racing events. Area measurements were made during race preparation, practice, qualification, and competition. Personal dosimetry measurements were conducted on drivers, crew members, infield staff, and spectators. Findings showed time-weighted averages (TWA) that ranged from 94 decibels A-weighted (dBA) for spectators to 114 dBA for car drivers. Peak sound-pressure levels exceeded the maximum allowable limit of 140 decibels (dB) during race competitions. Personal exposure measurements exceeded the NIOSH recommended exposure limit (REL) of 85 dBA as an 8-h TWA in less than a minute for one driver during practice, within 2 min

for pit crew and infield staff, and 7 to 10 min for spectators during the race. Hearing protection use was variable and intermittent among crew, staff, and spectators. Among drivers and crew, there was greater concern for communication performance than for hearing protection.

4:20

4pNSb8. Sports acoustics: Using sound from resonant shells, vibrating cylinders, strumming shafts, and water impact to evaluate athletic performance. David G. Browning (Dept. of Phys., Univ. of Rhode Island, Kingston, RI 02881, decibeldb@aol.com) and Peter M. Scheifele (Univ. of Connecticut, Storrs, CT 06269)

The sound from equipment used and/or specific acts during athletic competition, such as hitting a baseball with an aluminum bat, carries beyond the playing field and can provide a nonobtrusive method to evaluate athletic performance—such as where on the bat the ball was hit. Standardized equipment guarantees repeatability, for example, every volleyball resonates at the same frequency. Each major sport can have unique noise interference which in some circumstances can be overwhelming, and the distance from the sound source can vary significantly during a game. Still, it will be shown that useful performance information can be obtained under realistic conditions for at least the following sports: volleyball, softball, baseball, golf, swimming and diving, soccer, and football.

FRIDAY AFTERNOON, 1 DECEMBER 2006

WAIANAE ROOM, 1:30 TO 6:20 P.M.

Session 4pPA

Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration: Sound Propagation in Inhomogeneous Media II

Takahiko Otani, Cochair
Doshisha Univ., Lab. of Ultrasonic Electronics, Kyotonabe-shi, Kyoto 610-0321, Japan

James G. Miller, Cochair
Washington Univ., Dept. of Physics, 1 Brookings Dr., St. Louis, MO 63130

Invited Paper

1:30

4pPA1. Observables and prediction modeling in the presence of ultra-wideband heterogeneity. John J. McCoy (The Catholic Univ. of America, Washington, DC 20064)

Underlying virtually all propagation and scattering models is an intuitive understanding; the acoustic field is observable using a device of a sufficiently small size to obtain a sufficiently dense set of discrete measurements. This assumes the field variation cuts off at an inner length scale larger than the device size, assuring that no information is lost to the inherent spatial averaging in any measurement. This understanding is faulty in the presence of environment heterogeneity observed on an extreme range of length scales. The reason is that all physical devices have finite accuracy, which limits their ability to capture variation on scales significantly larger than their size, in the presence of variation on intermediate scales. A more refined understanding of the ability to observe a field requires multiple devices, an unbounded hierarchy in the limit, to obtain multiple dense sets of discrete "observables." This, then, suggests a different class of prediction models for environments with ultra-wideband heterogeneity, expressed in multiple sets of discrete variables, each set describing field variation in a limited subband. A framework for formulating these prediction models and their application to a scenario for which environment heterogeneity has no inner scale cutoff is presented.

Contributed Papers

4pPA2. Parameters arising from the Burridge-Keller formulation for poroelastic media, especially for granular media and marine sediments. Allan D. Pierce, William M. Carey, and Paul E. Barbone (Boston Univ., Boston, MA 02215)

It was previously shown [Pierce *et al.* (2006)] that the Burridge-Keller formulation [J. Acoust. Soc. Am. (1981)] rigorously justifies the low-frequency version [(1956a)] of Biot's equations. Implementation involves two microscale problems: (1) incompressible viscous flow driven in a highly irregular space with rigid walls by a uniformly and externally applied apparent pressure distribution and (2) elastostatic deformation of an intricate elastic web caused by the joint influence of a distributed constant body force and by uniform tractions (including an external pressure) on the web's exposed surface. Microscale averages produce the Biot "constants." Theoretical devices of applied mechanics and mathematics yield estimates of these and related parameters. In particular, it is shown that Wood's equation is a reasonable first approximation for the sound speed in sediments in the low-frequency limit. The formulation also yields an estimation for the sound speed in the high-frequency limit, when the viscous boundary layers become thin. The well-known result that the attenuation varies as $f^{1/2}$ in the high-frequency limit also results without the necessity of Biot's heuristic patching theory. Various heuristic approximations due to Gassmann, to Geertsma and Smit, and to Stoll and Bryant are analytically and numerically assessed.

4pPA3. Sound propagation in the mixtures of liquid and solid aggregate; similarities at micro- and nanoscales.. Hasson M. Tavossi (Dept. of Physical and Environ. Sci., Mesa State College, 1100 North Ave., Grand Junction, CO 81504)

Sound propagation phenomena in certain liquid and solid aggregate mixtures, at micrometer scales, in some cases resemble the wave propagation behaviors of materials observed at nanometer and atomic scales. For example, it can be shown that the sound wave dispersion, attenuation, and cutoff-frequency effects depend on the same structural parameters as those observed at nano or atomic levels and are similar at both scales. Therefore, to investigate theoretical models of wave and matter interactions it is more convenient to use, as experimental tools, the readily analyzable models of wave propagation, in mixtures of solid and liquid, constructed at micrometer scales. Theoretical findings on sound propagation in the mixtures of liquid and solid particles at micrometer scales will be discussed. These results show the resemblance to the behavior of acoustic phonons, the lattice thermal vibrations of crystalline structures, at radically different scales. Experimental data on wave dispersion, attenuation, bandpass, and cutoff frequency effects, measured for sound propagation, in inhomogeneous materials consisting of mixtures of solid and liquid will be presented, showing the similarities of wave propagation behaviors at micro- and nanoscales.

4pPA4. Nonlinear surface waves in soil. Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Nonlinear effects in surface waves propagating in soil are investigated theoretically. Analytic solutions are derived for the second harmonics and difference frequency waves generated by a bifrequency primary wave propagating at moderate amplitude. The soil is modeled as an isotropic solid. As such, its elastic properties are described by five elastic constants, two at second order in the strain energy density (the shear and bulk moduli) and three at third order. Nonlinear propagation of the surface waves is based on a theory developed previously [Zabolotskaya, J. Acoust. Soc. Am. **91**, 2569–2575 (1992)]. Elements of the nonlinearity matrix associated with the interacting spectral components are expressed in terms of the five elastic constants. It was found convenient to express the nonlinearity matrix for soil as a function of a nonlinearity parameter corresponding to B/A for liquids, particularly for saturated soils exhibiting liquidlike properties. This nonlinearity parameter can vary by several orders of magnitude. For soils with shear wave speeds less than 20% of the compressional wave speeds, the nonlinearity of surfaces waves is found to be independent of the third-order elastic constants and dependent only on the shear modulus. [Work supported by ONR.]

4pPA5. The measurement of the hysteretic nonlinearity parameter of a field soil by the phase shift method: A long-term survey. Zhiqu Lu (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, Univ., MS 38677)

Soil properties significantly affect the performance of the acoustic landmine detection. The climate and seasonal changes cause the variations of soil properties and smear landmine signature over time. On the other hand, soil is a complicated granular material that exhibits strong nonlinear acoustic behaviors. To understand the weather and seasonal effects on nonlinear acoustic behaviors of soils, a phase shift method is used to measure the hysteretic nonlinearity parameter of a field soil. The technique is based on measuring the variation of phase difference between two transducers, i.e., the phase shift, induced by changing sound level. The hysteretic nonlinear parameter can be extracted from the measured phase shift as a function of sound level or dynamic strain. In a long-term survey, the nonlinearity parameter, sound speed, and environmental conditions such as temperature, moisture, soil water potential, and rainfall precipitation are measured. It is found that the nonlinearity parameter is much more sensitive than sound speed to the climate change. Soil water potential is the predominant factor that affects the nonlinearity parameter and sound speed of the shallow field soil.

4pPA6. Nonlinear acoustic landmine detection: Comparison of soil nonlinearity with soil-interface nonlinearity. Murray S. Korman, Kathleen E. Pauls, Sean A. Genis (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402), and James M. Sabatier (Univ. of Mississippi, Univ., MS 38677)

To model the soil-top plate interface in nonlinear acoustic landmine detection, the soil-plate oscillator was developed [J. Acoust. Soc. Am. **116**, 3354–3369 (2004)]. A Lexan plate (2.39 mm thick, 18.5 cm diameter) is clamped at an inside diameter of 11.8 cm between two metal flanges. Dry sifted masonry sand (2-cm layer) is placed over the plate. Turning curves experiments are performed by driving a loudspeaker (located over the sand) by a swept sinusoid. The acceleration versus frequency is measured near resonance on a swept spectrum analyzer using an accelerometer centered on the surface. The corresponding backbone curve exhibits a linear decrease in resonant frequency f versus increasing acceleration, where $a = -a_o(f-f_o)/f_o$. Define a nonlinear parameter $\alpha = 1/a_o$. When the elastic plate is replaced by a "rigid" plate, α decreased from 0.128 to 0.070 (s^2/m), while f_o increased from 191 to 466 Hz. When a cylindrical drumlike mine simulant (rigid walls, thin acrylic top-plate) was buried 2 cm deep in a concrete sand box, "on the mine" results yielded $\alpha = 0.30$ (s^2/m) with $f_o = 147$ Hz, while "off the mine," $\alpha \sim 0.03$ (s^2/m) at $f_o = 147$ Hz. [Work supported by ONR.]

4pPA7. Causality conditions and signal propagation in bubbly water. Gregory J. Orris, Dalcio K. Dacol, and Michael Nicholas (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Acoustic propagation through subsurface bubble clouds in the ocean can exhibit signal travel times with enormous variations depending on the acoustic signal frequency, bubble size distribution, and void fraction. Recent theories have predicted large variations in phase speeds and attenuation that have been largely validated for frequencies well below and well above bubble resonance. However, great care must be exercised when

theoretically treating signal propagation at frequencies near resonance, termed the "Anomalous Absorption Regime" nearly 100 years ago in the pioneering work of Sommerfeld [A. Sommerfeld, Physik. Z. **8**, 841 (1975)] while investigating aspects of electromagnetic causality. We will discuss similarities between acoustic propagation in bubbly media and electromagnetic propagation in the presence of a conducting medium. We show that the signal travel time is dependent on the behavior of the dispersion formula in the complex frequency plane and place limits on the range of validity of these formulas, leading naturally to the necessary modifications to the current dispersion formulas to bring them into compliance with causality. Finally, we present theoretical results for the velocity of signals for a representative environment of experimental work carried out at the Naval Research Laboratory. [Work supported by the ONR.]

3:20

4pPA8. Measurements of the attenuation and sound speed in bubbly salt water. Gregory J. Orris, Dalcio K. Dacol, and Michael Nicholas (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Bubble clouds were injected from below the surface of a 144-cubic-meter water tank, wherein hydrophones were placed at varying distances from an acoustic source. Measurements were made over a wide range of frequencies to verify and validate the theoretical predictions of the relevant dispersion formula. This work was undertaken under a variety of conditions by varying the relevant environmental parameters: void fraction, temperature, and salinity. Void fractions were varied from roughly 0.02% to 0.1%. Temperatures ranged from 9 °C to 18 °C, and the salinity was varied from zero to approximately 10% of typical oceanic values. Particular attention was paid to tracking the phase of the transmitted signal as the frequency progressed toward resonance starting from 100 kHz. This yielded phase-speed measurements in an essentially free-field environment using a modified version of phase spectral analysis. Time-of-flight measurements gave signal velocities, while the received energy yielded the attenuation. Results are compared to theoretical calculations, leading to the conclusion that current theoretical dispersion formula requires modification. [This work supported by the ONR.]

3:35

4pPA9. Efficient computation of 3-D acoustical scattering from multiple arbitrarily shaped objects using the boundary element method/fast multipole method (BEM/FMM). Nail A. Gumerov and Ramani Duraiswami (Perceptual Interfaces and Reality Lab., Inst. for Adv. Comput. Studies, Univ. of Maryland, College Park, MD 20742)

Many applications require computation of acoustic fields in systems consisting of a large number of scatterers, which may have complex shape. Despite the boundary element method being a well-known technique for solution of the boundary value problems for the Helmholtz equation, its capabilities are usually limited by the memory and speed of computers, and conventional methods can be applicable to relatively small (up to order of 10 000 boundary elements) problems. We developed and implemented an efficient computational technique, based on an iterative solver employing generalized minimal residual method in combination with matrix-vector multiplication speeded up with the fast multipole method. We demonstrate that this technique has $O(N)$ memory and computational complexity and enables solution of problems with thousands of scatterers (millions of boundary elements) on a desktop PC. The test problems solved are of moderate frequency (up to $kD \leqslant 150$, where k is the wavenumber and D is the size of the computational domain). Solution of large scale scattering problems was tested by comparison with the FMM-based T-matrix method applicable for simple shape objects reported earlier [Gumerov and Duraiswami, J. Acoust. Soc. Am., **117**(4), 1744–1761 (2005)], visualization, and physical interpretation of the results.

3:50

4pPA10. Fast acoustic integral-equation solver for complex inhomogeneous media. Elizabeth Bleszynski, Marek Bleszynski, and Thomas Jaroszewicz (Monopole Res., 739 Calle Sequoia, Thousand Oaks, CA 91360, elizabeth@monopoleresearch.com)

We describe elements and representative applications of an integral-equation solver for large-scale computations in acoustic wave propagation problems. In the solver construction we used elements of our previously developed fast integral-equation solver for Maxwell's equations. In comparison with the conventional integral equation approach (method of moments), our solver achieves significant reduction of execution time and memory through the FFT-based matrix compression. One particular aspect of the solver we discuss, pertinent to its high efficiency and accuracy, is an efficient treatment of problems associated with subwavelength discretization. We illustrate the approach and its application on the example of a numerical simulation of acoustic wave propagation through the human head. [Work was supported by a grant from AFOSR.]

4:05

4pPA11. Models for acoustic scattering in high contrast media. Max Denis, Charles Thompson, and Kavitha Chandra (Univ. Massachusetts Lowell, One University Ave., Lowell, MA 01854)

In this work a numerical method for evaluating backscatter from a three-dimensional medium having high acoustic contrast is presented. The solution is sought in terms of a perturbation expansion in the contrast amplitude. It is shown that limitations of the regular perturbation expansion can be overcome by recasting the perturbation sequence as a rational fraction using Padé approximants. The resulting solution allows for an accurate representation of the pressure and allows for the poles in the frequency response to be modeled. The determination of the pulse-echo response for a high-contrast medium is discussed and presented.

4:20

4pPA12. Multiple scattering and visco-thermal effects. Aroune Duclos, Denis Lafarge, Vincent Pagneux (Laboratoire d'Acoustique de l'Universite du Maine, Ave. Olivier Messiaen, 72085 Le Mans, France), and Andrea Cortis (Lawrence Berkeley Natl. Lab., Berkeley, CA 94720)

For modeling sound propagation in a rigid-framed fluid-saturated porous material it is customary to use frequency-dependent density and compressibility functions. These functions, which describe "temporal" dispersion effects due to inertial/viscous and thermal effects, can be computed by FEM in simple geometries and give complete information about the long-wavelength properties of the medium. When the wavelength is reduced, new effects due to scattering must be considered. To study this, we consider solving the sound propagation problem in a 2-D "phononic crystal" made of an infinite square lattice of solid cylinders embedded in a fluid. An exact multiple-scattering solution is first developed for an ideal saturating fluid and then generalized to the case of visco-thermal fluid, by using the concept of visco-thermal admittances. The condition to use this concept is that the viscous and thermal penetration depths are small compared to the cylinder radius. We validate our results in the long-wavelength regime by direct comparisons with FEM data [A. Cortis, "Dynamic parameters of porous media," Ph.D. dissertation (Delft U.P., Delft, (2002)]. When frequency increases, differences appear between the long-wavelength solution and the exact multiple-scattering solution, which could be interpreted in terms of "spatial" dispersion effects.

4:35

4pPA13. Effective parameters of periodic and random distributions of rigid cylinders in air. Daniel Torrent and José Sánchez-Dehesa (Wave Phenomena Group, Nanophotonics Technol. Ctr., Polytechnic Univ. of Valencia, C/Camino de vera s/n., E-46022 Valencia, Spain)

The scattering of sound by finite-size clusters consisting of two-dimensional distributions (periodic and random) of rigid cylinders in air is theoretically studied in the low-frequency limit (homogenization). Analyti-

cal expressions for the effective density and sound speed obtained in the framework of multiple scattering will be reported. For the case of circular-shaped cluster, we have theoretically analyzed the homogenization as a function of the filling fraction, the type of arrangement of the cylinders in the cluster (hexagonal and square lattice), and the number of cylinders in the cluster. When the number of cylinders in the cluster is small we found that for certain ''magic numbers'' their effective parameters (sound speed and density) are the same as those of the corresponding infinite array. [Work supported by MEC of Spain.]

4:50

4pPA14. The application of *k*-space acoustic propagation models to biomedical photoacoustics. Benjamin T. Cox (Dept. of Med. Phys. and Bioengineering, Univ. College London, Gower St., London, WC1E 6BT, UK, bencox@mpb.ucl.ac.uk), Simon. R. Arridge, and Paul C. Beard (Univ. College London, Gower St., London, WC1E 6BT, UK)

k−space models for broadband acoustic pulse propagation differ from pseudo−spectral time domain (PSTD) models in their treatment of the time step. By replacing a finite−difference scheme with a propagator, exact for homogeneous media, larger time steps can be taken without loss of accuracy or stability and without introducing dispersion. Three k−space models for modeling photoacoustically generated (PA) pulses are described here. A very simple, exact, model of PA propagation in a homogeneous fluid is used to introduce the k−space propagator, and two models of propagation in heterogeneous media, originally designed for modeling scattering in soft tissue, are adapted for use in photoacoustics [Mast *et al.*, IEEE Trans. UFFC **48**, 341−354 (2001); Tabei *et al.*, J. Acoust. Soc. Am. **111**, 53−63 (2002)]. Our motivation for describing these models comes from biomedical PA imaging, in which one of the current limitations is the assumption that soft tissue has a uniform sound speed. Efficient, accurate, and simple−to−encode forward models such as these are very useful for studying the effects of the heterogeneities encountered in practice. They may also be useful in designing PA imaging schemes that can account for acoustic heterogeneities. [This work was funded by the EPSRC, UK]

5:05

4pPA15. Emergence of the acoustic Green's function from thermal noise. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA, Earth System Res. Lab., 325 Broadway, Boulder, CO 80305, oleg.godin@noaa.gov)

Recently proposed applications of noise cross-correlation measurements to passive remote sensing range from ultrasonics and acoustic oceanography to helioseismology and geophysics, at wave frequencies that differ by more than ten orders of magnitude. At the heart of these applications is the possibility to retrieve an estimate of a deterministic Green's function from long-range correlations of diffuse noise fields. Apparently, S. M. Rytov [*A Theory of Electrical Fluctuations and Thermal Radiation* (USSR Academy of Sciences, Moscow, 1953)] was the first to establish theoretically a simple relation between the Green's function and the two-point correlation function of fluctuations of wave fields generated by random sources. He used reciprocity considerations to analyze fluctuations of electromagnetic fields. In this paper, an acoustic counterpart of Rytov's approach is applied to derive exact and asymptotic relations between respective acoustic Green's functions and cross-correlation of thermal noise in inhomogeneous fluid, solid, and fluid-solid media. Parameters of the media are assumed to be time independent, but can be arbitrary functions of spatial coordinates. Theoretical results obtained are compared to those previously reported in the literature.

5:20

4pPA16. Simulation of elastic wave scattering in living tissue at the cellular level. Timothy E. Doyle and Keith H. Warnick (Dept. of Phys., Utah State Univ., 4415 Old Main Hill, Logan, UT 84322-4415, timdoyle@cc.usu.edu)

Elastic wave scattering in biological tissue has been simulated at the cellular level by incorporating a first-order approximation of the cell structure and multiple scattering between cells. The cells were modeled with a concentric spherical shell-core structure embedded in a medium, with the core, shell, and medium representing the cell nucleus, the cell cytoplasm, and the extracellular matrix, respectively. Using vector multipole expansions and boundary conditions, scattering solutions were derived for a single cell with either solid or fluid properties for each of the cell components. Multiple scattering between cells was simulated using addition theorems to translate the multipole fields from cell to cell and using an iterative process to refine the scattering solutions. Backscattering simulations of single cells demonstrated that changes in the nuclear diameter had the greatest effect on the frequency spectra as compared to changes in cell size, density, and shear modulus. Wave field images and spectra from clusters of up to several hundred cells were also simulated, and they exhibited phenomena such as wave field enhancement at the cell membrane and nuclear envelope due to the scattering processes. Relevant applications for these models include ultrasonic tissue characterization and ultrasound-mediated gene transfection and drug delivery.

5:35

4pPA17. Acoustic analog of electronic Bloch oscillations and Zener tunneling. José Sánchez-Dehesa, Helios Sanchis-Alepuz, Yu. A. Kosevich, and Daniel Torrent (Wave Phenomena Group, Polytechnic Univ. of Valencia, C/Camino de Vera s.n., E-46022 Valencia, Spain)

The observation of Bloch oscillations in sound propagation through a multilayer of two different fluidlike components is predicted. In order to obtain the equivalent to the acoustic analog of a Wannier-Stark ladder [E. E. Mendez *et al.*, Phys. Rev. Lett. **60**, 2426–2429 (1988)], a set of cavities with increasing thickness is employed. Bloch oscillations were theoretically predicted as time-resolved oscillations in transmission in direct analogy to electronic Bloch oscillations in semiconductor superlattices [J. Feldmann *et al.*, Phys. Rev. B **46**, R7252–R7255 (1992)]. Finally, an experimental setup is proposed to observe the phenomenon by using arrays of cylindrical rods in air, which acoustically behaves as a fluidlike system with effective sound velocity and density [D. Torrent *et al.*, Phys. Rev. Lett. **96**, 204302 (2006)]. For the proposed system, Bloch oscillations and Zener tunneling are confirmed by using multiple scattering simulations. [Work supported by MEC of Spain.]

5:50

4pPA18. Comparison of time reversal acoustic and prefiltering methods of focusing of tone burst signals. Bok Kyoung Choi (Korea Ocean Res. and Development Inst., Sangrok-gu, 426-744, Korea), Alexander Sutin, and Armen Sarvazyan (Artann Labs., Inc., West Trenton, NJ 08618)

The concept of time reversal acoustics (TRA) provides an elegant possibility of both temporal and spatial concentration of acoustic energy in highly inhomogeneous media. TRA-based focusing is typically used for generation of short acoustic pulses, however, in some medical and industrial applications, longer pulses are required. TRA focusing of longer signals leads to an increase of side lobes in temporal and spatial domains. Another method for focusing, known as prefiltering, is based on measurements of the impulse response, which relates the signal at the TRA transmitter to that at the focusing point. After evaluating the impulse response, the excitation signal may be calculated to generate the desired waveform in the focus point. This method allows signal generation with any desired form including long tone-burst signals. Experiments on comparison TRA and prefiltering methods of ultrasound focusing were conducted in the frequency band of 200–1000 kHz. In the experiments, focused acoustic pulses with various forms and duration were generated: triangular, rectan-

4p FRI. PM

gular, and amplitude-modulated tone burst signals. The prefiltering modes provide better temporal compression of the focused signal, and the signal energy outside the main pulse in the prefiltering mode was shown to be much lower than that in standard TRA focusing.

6:05

4pPA19. Modeling quasi-one-dimensional sound propagation in ducts having two propagation media using a cross-sectional averaging theory. Donald Bliss and Lisa Burton (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708)

Sound propagation of quasi-one-dimensional waves through a uniform duct partially filled with porous material has been studied theoretically and experimentally. The porous material makes the effective propagation wave number in the duct complex. A fairly simple theory based on cross-sectional averaging is derived and tested and found to work extremely well up to fairly high frequency. Interestingly, the basic theory depends only on the ratio of cross-sectional areas and the properties of the individual propagation media, but not on the specific configuration of material in a cross section. A higher order correction is developed to achieve excellent accuracy to very high frequency. This correction includes a coefficient that does depend on the specific cross-sectional configuration. Results are compared to exact solutions for layered and annular configurations, and also to experimental measurements with open cell foam as the porous material. An interesting application is to use measured wave numbers to predict the complex effective density and sound speed of porous media samples partially filling the duct. Other applications include fairly simple improved predictions of the behavior of sound in ducts lined with, or partially filled with, bulk reacting absorbing material.

FRIDAY AFTERNOON, 1 DECEMBER 2006

WAIALUA ROOM, 1:30 TO 4:20 P.M.

Session 4pPP

Psychological and Physiological Acoustics: Auditory Physiology

G. Christopher Stecker, Cochair

Univ. of Washington, Dept. of Speech and Hearing Science, 1417 NE 42nd St., Seattle, WA 98105

Shigeto Furukawa, Cochair

NTT Communication Science Labs., Human and Information Science Lab., 3-1 Morinosato-wakamiya, Atsugi-shi, Kanagawa-ken 243-0198, Japan

Chair's Introduction—1:30

Contributed Papers

1:35

4pPP1. Transmission of bone-conducted sound measured acoustically and psycho-acoustically. Sabine Reinfeldt (Signals & Systems, Chalmers Univ. of Technol., SE-412 96 Goteborg, Sweden), Stefan Stenfelt (Linkoping Univ., SE-581 83 Linkoping, Sweden), and Bo Hakansson (Chalmers Univ. of Technol., SE-412 96 Goteborg, Sweden)

The transcranial transmission is important in the bone-conducted (BC) audiometry where the BC hearing thresholds depend on the stimulation position. It is also important for fitting of BC hearing aids; the transcranial transmission determines the amount of the sound that reaches the contralateral cochlea. Previous reported transcranial transmission results seem to depend on the method used. Here, a comparison between the transcranial transmission measured with BC hearing thresholds and ECSP is performed for both open and occluded ear canal. A BC transducer provided stimulation at both mastoids and the forehead. The ECSP was measured with a probe microphone and the BC hearing thresholds were obtained while masking the nontest ear. The transcranial transmission was determined as the BC hearing threshold or the ECSP for contralateral stimulation relative ipsilateral stimulation. The transmission from the forehead was calculated in a similar way. The transcranial transmission was similar for BC hearing thresholds and ECSP above 800 Hz; this indicates that the ECSP can be used as an estimator of the relative hearing perception by BC. The transcranial transmission results are also similar to vibration measurements of the cochleae made in earlier studies. Hence, vibration measurements of the cochleae can also estimate relative BC hearing.

1:50

4pPP2. Customization of head-related transfer functions using principal components analysis in the time domain. Ki H. Shin and Youngjin Park (Dept. of Mech. Eng., Korea Adv. Inst. of Sci. and Technol. (KAIST), Sci.e Town, Daejeon, 305-701, Republic of Korea)

Pinna responses were separated from HRIRs (head-related impulse responses) of 45 subjects in the CIPIC HRTF (head-related transfer function) database and modeled as linear combinations of five basic temporal shapes (basis functions) by PCA (principal components analysis) accounting for more than 90% of the variance in the original pinna responses per each selected elevation angle in the median plane. By adjusting the weight of each basis function computed for a specific height to replace the pinna response in the KEMAR HRIR at the same height with the resulting pinna response and listening to the filtered stimuli over headphones, four subjects were able to create a set of median HRIRs that outperformed the KEMAR HRIRs in producing elevation effects in the median plane. Since the monoaural spectral features due to the pinna are strongly dependent on elevation instead of azimuth, similar elevation effects could also be generated at different azimuthal positions simply by inserting the customized median pinna responses into the KEMAR HRIRs at other azimuths and varying the ITD (interaural time difference) according to the direction as well as the size of the subject's own head.

4pPP3. An electrophysiological measure of basilar membrane nonlinearity in humans. Christopher J. Plack (Dept. of Psych., Lancaster Univ., Lancaster, LA1 4YF, England) and Ananthanarayan Krishnan (Purdue Univ., West Lafayette, IN 47906)

A behavioral measure of the basilar membrane response can be obtained by comparing the growth in forward masking for maskers at, and well below, the signal frequency. Since the off-frequency masker is assumed to be processed linearly at the signal place, the difference in masking growth with level is thought to reflect the compressive response to the on-frequency masker. The present experiment used an electrophysiological analog of this technique, based on measurements of the latency of wave V of the auditory brainstem response elicited by a 4-kHz, 4-ms pure tone, presented at 65 dB SPL. Responses were obtained in quiet and in the presence of either an on-frequency or an off-frequency (1.8 kHz) pure-tone forward masker. Wave V latency increased with masker level, although the increase was greater for the off-frequency masker than for the on-frequency masker, consistent with a more compressive response to the latter. Response functions generated from the data showed the characteristic shape, with a nearly linear response at lower levels, and 5:1 compression at higher levels. However, the breakpoint between the linear region and the compressive region was at about 60 dB SPL, higher than expected on the basis of previous physiological and psychophysical measures.

2:20

4pPP4. Possible involvement of the spiral limbus in chinchilla cochlear mechanics. William S. Rhode (Dept. of Physiol., 1300 University Ave., Madison, WI 53706, rhode@physiology.wisc.edu)

Differences between cochlear mechanical tuning curves and those of auditory nerve fibers (ANFs) exist. In particular, mechanical transfer functions exhibit a high-frequency plateau; ANFs frequency threshold curves (FTCs) do not. ANF-FTCs may have a low-frequency slope due to a velocity forcing function operating on inner hair cells at low frequencies. Neither basilar membrane velocity nor displacement adequately explain the entire ANF tuning curve. A displacement sensitive interferometer was used to study basilar membrane and spiral limbus mechanics in the 6-kHz region of the chinchilla cochlea. The spiral limbus vibrates at the same phase as the basilar membrane nearly up to the location's characteristic frequency. In the plateau region, the limbus appears to vibrate 0 to 20 dB less than the basilar membrane. The basilar membrane/limbus amplitude transfer function has a low-frequency slope of ~ 3 dB/oct at low frequencies and is ~ 10 dB lower than the basilar membrane amplitude at 1 kHz. It appears that spiral limbus vibration may contribute to the excitation of the cilia of the inner hair cells. [Work supported by NIDCD grant R01 DC001910.]

2:35

4pPP5. The effects of antioxidants on cochlear mechanics. Barbara Acker-Mills, Melinda Hill, and Angeline Ebuen (U.S. Army Aeromedical Res. Lab., P.O. Box 620577, Fort Rucker, AL 36362-0577, barbara.acker@us.army.mil)

Several studies have evaluated the effect of N-acetylcysteine (NAC) on temporary threshold shifts (TTSs) in humans. Work at USAARL found that NAC did not reduce TTSs compared to a placebo, but suppressed otoacoustic emissions (OAEs) more than a placebo, indicating that NAC may reduce outer hair cell activity. Kramer *et al.* [JAAA, **17**(4), (2006)] found similar results, where NAC did not affect thresholds in people who had TTSs from exposure to loud music. However, OAEs also did not differ between NAC and placebo. Toppilla *et al.* [XXII Barany Society Meeting (2002)] measured thresholds and balance in people exposed to loud music and found that while NAC did not affect TTS, it reduced noise-induced balance impairment. The current study administered NAC and measured cochlear microphonics, compound action potentials, and OAEs to evaluate cochlear function. The vestibular myogenic potential was measured to assess the effect of NAC on the saccule. The results provide a comprehensive analysis of the effect of NAC on the auditory

system and one component of the vestibular system. [Work supported by the U.S. Army ILIR program. The work is that of the authors and is not necessarily endorsed by the U.S. Army or the Department of Defense.]

2:50–3:05 Break

3:05

4pPP6. Time characteristics of distortion product otoacoustic emissions recovery function after moderate sound exposure. Miguel Angel Aranda de Toro, Rodrigo Ordoñez, and Dorte Hammershøi (Dept. of Acoust., Aalborg Univ., Fredrik Bajers Vej 7 B5, DK-9220, Aalborg, Denmark, maat@acoustics.aau.dk)

Exposure to sound of moderate level temporarily attenuates the amplitude of distortion product otoacoustic emissions (DPOAEs). These changes are similar to the changes observed in absolute hearing thresholds after similar sound exposures. To be able to assess changes over time across a broad frequency range, a detailed model of the recovery time characteristics is necessary. In the present study, the methodological aspects needed in order to monitor changes in DPOAEs from human subjects measured with high time resolution are presented. The issues treated are (1) time resolution of the measurements, (2) number of frequency points required, and (3) effects in fine structures, are they affected with the exposure? [Work supported by the Danish Research Council for Technology and Production.]

3:20

4pPP7. Probability characteristics of neural coincidence detectors in the brainstem. Ram Krips and Miriam Furst (Dept. of Elec. Eng. Systems, Faculty of Eng., Tel Aviv Univ., Tel Aviv 69978, Israel, mira@eng.tau.ac.il)

Auditory neural activity in the periphery is usually described as non-homogeneous Poisson process (NHPP), characterized as either EE or EI, which operates as a coincidence detector. In order to apply a general probabilistic analysis of those brainstem nuclei activity, the stochastic properties of the axons that exit the EE and EI nuclei is essential. An analytical analysis of the probability characteristics of the output of an EE nucleus (EEout) will be presented. Assuming that an EE nucleus receives inputs from two neurons, each behaves as an NHPP with instantaneous rates λ_1 and λ_2, and an output spike is generated if both spike at a coincidence window (Δ) which is relatively small (this matches biological findings). Then EEout is also a NHPP with instantaneous rate $r(t) = \lambda_1(t)\int_{t-\Delta}^{t}\lambda_2(\zeta)\,d\zeta + \lambda_2(t)\int_{t-\Delta}^{t}\lambda_1(\zeta)\,d\zeta$. A similar derivation was applied for an EI nucleus. We found also that the output activity is NHPP for a relatively small coincidence window (Δ). The obtained IR is $r(t) = \lambda_e(t)[1 - \int_{t-\Delta}^{t}\lambda_i(\zeta)\,d\zeta]$, where λ_E and λ_I are the excitatory and inhibitory input IRs. On the other hand, for larger Δ, the output activity is not a Poisson process. Those derivations enable theoretical analysis and performance evaluation of large neural networks, which perform binaural tasks as ITD and ILD.

3:35

4pPP8. Musical expertise and concurrent sound segregation. Benjamin Rich Zendel and Claude Alain (Rotman Res. Inst., Baycrest Ctr. & Dept. of Psych., Univ. of Toronto, Toronto, ON M6A 2E1, Canada)

There is growing evidence suggesting that musical training can improve performance in various auditory perceptual tasks. These improvements can be paralleled by changes in scalp recorded auditory evoked potentials (AEPs). The present study examined whether musical training modulates the ability to segregate concurrent auditory objects using behavioral measures and AEPs. Expert musicians and nonmusicians were presented with complex sounds comprised of six harmonics (220, 440, 660 Hz, etc.). The third harmonic was either tuned or mistuned by 1%–16% of its original value. Mistuning a component of a harmonic complex results in the precept of a second auditory object. Stimuli were presented passively (no response) and actively (participants responded by indicating if they heard one sound or two sounds). Behaviorally both musicians and

4p FRI. PM

nonmusicians perceived a second auditory object at similar levels of mistuning. In both groups, complex sounds generated N1 and P2 waves at fronto-central scalp regions. The perception of concurrent auditory objects was paralleled by an increased negativity around 150 ms post-stimulus onset. This increased negativity is referred to as object-related negativity (ORN). Small differences between musicians and nonmusicians were noted in the ORN. The implication of these results will be discussed in terms of current auditory scene analysis theory.

3:50

4pPP9. Study of auditory temporal processing of language and music in stroke patients. Li Hsieh and Tamara Baubie (Dept. of Commun. Sci. and Disord., Wayne State Univ., 207 Rackham Hall, 60 Farnsworth, Detroit, MI 48202)

This study focused on functional lateralization of temporal processing of language and music for nine stroke patients and nine normal controls. Subjects were asked to discriminate short versus long sounds in ABX tasks, and then to reproduce sequences of short and long sounds they heard. The reproduction tasks consisted of sequences of 3, 4, and 5 sounds. The linguistic stimuli were nonsense CVC syllables, and the musical stimuli were computer-generated musical notes with the timbres of French horn and trombone. Both linguistic and music stimuli were controlled for frequency and intensity, and varied only for duration (i.e., 500 and 750 ms). Our findings are consistent with previous studies; the left hemisphere specializes in linguistics, while the right hemisphere in music. Moreover, both hemispheres appeared to work closely together in processing temporal information. Both left- and right-hemisphere-damaged patients performed worse than controls. Most subjects performed better with music than language. Patients with left posterior lesions performed worse compared to patients with left or right anterior lesions, particularly with linguistic stimuli. Patients with right anterior lesions not only involved with temporal processing in music, but also in linguistic stimuli. Our study provided additional information regarding transient temporal processing in language and music.

4:05

4pPP10. Human bioacoustic biology: Acoustically anomalous vocal patterns used to detect biometric expressions relating to structural integrity and states of health. Sharry Edwards (Sound Health Res. Inst., P.O. Box 267, Sauber Res. Ctr., Albany, OH 45710)

Computerized analyses of acoustically anomalous vocal patterns are being used as biomarkers for predictive, prediagnostic, and efficient management of individual biological form and function. To date, biometrically distinct vocal data have resulted in outcomes that would be considered improbable by contemporary medical standards. For instance, independent EMG conclusions confirmed the regeneration of nerve tissue for a multiple sclerosis patient who used acoustic bioinformation to guide his primary therapy. Another study monitored the amounts of synthetic labor hormones present during induced labor. False labor costs amount to millions of dollars each year in insurance and hospital resources. The use of noninvasive, possibly remote, vocal profiling could ameliorate such costs. Anomalous vocal acoustics are being investigated by many health-related organizations including Pfizer Pharmaceuticals and the Institute of Automatic Control Engineering in Taiwan. Complementary research studying molecular frequencies of cellular chemistry is being done by James Gimjewski, Ph.D., UCLA, Department of Chemistry and Biochemistry. Known as BioAcoustic Biology, this research modality has the potential to advance current health care standards for biological function, disease processes, and metabolism. Organizations such as the Acoustical Society of America are considering standards for technically defining human bioacoustics. This paper would expand that language.

FRIDAY AFTERNOON, 1 DECEMBER 2006

HONOLULU ROOM, 1:00 TO 3:20 P.M.

Session 4pSAa

Structural Acoustics and Vibration: Vehicle Interior Noise and Vibration

Courtney B. Burroughs, Cochair
Noise Control Engineering, Inc. 1241 Smithfield St., State College, PA 16801

Hiroaki Morimura, Cochair
Japan Science and Technology Agency, Hosei Univ., Dept. of Mechanical Engineering, 3-7-2 Kajino-cho, Kogamei-city, Tokyo, Japan

Invited Papers

1:00

4pSAa1. Analytical model for flow-excited interior cavity resonance and its application to the Stratospheric Observatory for Infrared Astronomy (SOFIA). Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The Stratospheric Observatory for Infrared Astronomy (SOFIA) is a joint effort between NASA and the German Space Agency that has installed a 20 000-kg telescope in a modified 747-SP. The modifications entailed constructing bulkheads, one of which is used to provide the active mount for the telescope, and a door that rotates to open as much as one-quarter of the fuselage circumference to the atmosphere. This configuration represents a large compartment that can exhibit acoustic resonances at low frequencies. Concern arose that a Rossiter mode, which is an aerodynamic resonance in which vortices shed from the leading edge of a gap form a coherent standing pattern at certain speeds, would create a strong acoustic source for acoustic and structural modes, whose frequencies might coincide. A model consisting of a two-dimensional hard-walled waveguide having a Rossiter mode source and an elastic plate at one end was analyzed in order to understand these issues. Unlike conventional analyses of interior cabin noise, in which vibrating walls are the acoustic source, the elastic plate represents a compliant boundary that couples with the acoustic modes. The results lead to some general insights to the possibility of a ''triple resonance,'' as well as the role of structural compliance for cavities that are excited by turbulent external flows.

4pSAa2. Energy finite energy analysis for shipboard noise. Raymond Fischer, Leo Boroditsky (Noise Control Eng. Inc., 799 Middlesex Tnpk, Billerica, MA 01821), Layton Gilfroy (DRDC-Atlantic), and David Brennan (Martec Ltd.)

Machinery-induced habitability noise is difficult to model efficiently and accurately. The potential of energy finite-element analysis (EFEA) is compared to other prediction tools such as statistical energy analysis (SEA). This paper will explore the benefits and costs of EFEA with respect to SEA for acoustic modeling. The focus will be on issues relating to structural modeling for EFEA purposes. EFEA techniques will be evaluated to see if they possess the capabilities of verified SEA approaches for predicting habitability and radiated noise, where it is necessary to account for the impact of diverse marine constructions and sources such as the lack of machinery source information with respect to force or moment inputs or the finite impedance of machinery foundations. The effort proposed herein will provide the necessary engineering to research and identify salient features of EFEA that are potentially applicable for the detailed analysis of the acoustic environment and response of surface ships to various excitation sources. The paper will also address the pros and cons of SEA versus energy-finite element analysis (EFEA) methods used to predict the habitability noise of surface ship platforms. [This work is supported by an Office of Navy Research contract.]

4pSAa3. Spectral-based multicoordinate substructuring model for vehicle noise, vibration, and harshness refinement. Teik C. Lim (Mech., Industrial and Nuclear Eng., 598 Rhodes Hall, Box 210072, Univ. of Cincinnati, Cincinnati, OH 45221, teik.lim@uc.edu)

The success of vehicle NVH (noise, vibration, and harshness) refinement often depends on the ability to identify and understand noise and vibration transmission paths within the mid-frequency range, i.e., 200–1000 Hz, throughout the assembled structure. Due to the complexity of the dynamics in this frequency range, most modal or finite element-based methods do not possess the fidelity needed. To address this gap, a multicoordinate substructuring theory applying measured structural-acoustic and vibration spectra is applied. Three forms of substructuring formulation, namely the nondiagonal, block-diagonal, and purely diagonal coupling cases, are developed. The performances of these approaches are studied numerically, and the net effects of these coupling formulations on the predicted joint and free substructure dynamic characteristics, and system response, are determined. Conditions for applying the simpler coupler that can simplify the testing process and overcome computational deficiencies are also derived. When the measured data is noise contaminated, the singular value decomposition (SVD) algorithm is found to be quite helpful. Using an actual vehicle, a comprehensive analysis of the measured and predicted vehicle system responses is performed. The results are employed to develop an understanding of the primary controlling factors and transfer paths and to cascade system requirements to the substructure level.

4pSAa4. Practical application of digital simulation with physical test in vehicle virtual. Toshiro Abe (ESTECH Corp., 89-1 Yamashita-cho, Naka-ku, Yokohama-shi, Kanagawa-ken, Japan 231-0023, toshiro.abe@estech.co.jp)

In the current vehicle design and development program, the Virtual Product Development process (hereafter, VPD process) is the innovation for automotive industry, which improves product quality and shortens time to market. In general, valid CAE applications are the key component of the VPD process as well as physical tests being indispensable to create valid simulation technologies. This presentation explains how physical-test-based CAE leverages the VPD process. In particular, the presentation is based on how physical testing supports the VPD process and why the ESTECH philosophy is that "The essence of CAE lies in its synergy with Testing;" a philosophy that differentiates the company from the competition. To demonstrate these activities, case studies based on automotive dynamic and real time simulation will be presented. In the case studies, vehicle body NVH and brake noise analysis will be used to show the interaction between physical test and computer simulation. Finally, practical application of the VPD process in one of the leading Japanese automotive companies will be shown, where the effectiveness of the front loading in the actual vehicle development program and the actual deployment of VPD process to the Functional Digital Vehicle project in the powertrain design are to be presented.

Contributed Papers

4pSAa5. Active noise control simulations using minimization of energy density in a mock helicopter cabin. Jared Thomas, Stephan P. Lovstedt, Jonathan Blotter, Scott D. Sommerfeldt (Brigham Young Univ., Provo, UT 84602), Norman S. Serrano, and Allan Egelston (Silver State Helicopters, Provo, UT 84601)

Helicopter cabin noise is dominated by low-frequency tonal noise, making it an ideal candidate for active noise control. Previous work in active control of cabin noise suggests an energy density approach to be a good solution [B. Faber and S.D. Sommerfeldt, Global Control in a Mock Tractor Cabin Using Energy Density, Proc. ACTIVE 04, Sept. 2004.] Simulations for active noise control using energy density minimization have been made using recorded data from a Robinson R44 helicopter. Initial computer models show substantial noise reductions in excess of 6 dB at the error sensor are possible. Performance results for computer models and simulations in a mock cab for different control arrangements will be compared.

4pSAa6. Modeling airborne interior noise in full vehicles using statistical energy analysis. Arnaud Charpentier and Phil Shorter (ESI, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130, arnaud.charpentier@esi-group-na.com)

SEA is particularly well suited for predicting airborne noise in vehicles. The acoustic sources found in such environment are typically spatially distributed around the vehicle and can be well represented with SEA diffuse acoustic loads. Multiple transmission paths contribute to interior noise levels including (1) mass law transmission through trimmed panels, (2) resonant radiation from vibrating structures, and (3) flanking paths through gaskets, seals, and holes. All these transmission mechanisms may be modeled using SEA techniques. Finally, interior trim (including carpet, headliner, seats) is a key contributor to the acoustic performance of modern vehicles. The vehicle sound package has a significant impact on both the strength of the transmissions paths into the vehicle as well as the acoustic absorption in the cabin. Both these effects can be accounted for with SEA through detailed models of the trim. SEA models of full vehicles

are usually validated against experimental results at both component and system levels. The models can then be confidently used to (a) rank key design parameters governing interior levels and (b) quickly evaluate the impact of potential design changes. Example vehicle models and correlation results are presented here.

2:50

4pSAa7. Radiation from vibrating panels at high frequency including an inquiry into the role of edges and drive points. Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708)

In the high frequency limit, a vibrating finite panel is shown to have broadband power and directivity characteristics that can expressed analytically by a limited set of parameters. Two-dimensional problems with subsonic structural wave speed are considered. Three basic directivity patterns are identified, associated with right and left traveling waves and the correlation between them. The role of boundary conditions at the panel edge is illustrated, as are the effects of types of forcing. Overall, relatively simple broadband behaviors are revealed. The analytical characterization of the radiation is shown to be particularly straightforward in the high frequency broadband limit. Interestingly, the radiated mean-square pressures are independent of the panel length, indicating the radiation is associated with the edges and the drive point. However, the radiation patterns cannot be explained in terms of simple volumetric sources placed just at the edges and the drive point, showing that the often-stated idea of uncan-

celed volumetric sources at these locations is not correct except under very restricted circumstances. A correct physical interpretation of the radiation is provided both in physical space and in terms of spatial Fourier transforms.

3:05

4pSAa8. Some effect of trim and body panels on the low-frequency interior noise in vehicles. Andrzej Pietrzyk (Volvo Car Corp., Noise and Vib. Ctr., 405 31 Gothenburg, Sweden)

Structure borne noise dominates the interior noise in vehicles at low frequencies. One of the basic vinroacoustic characteristics of the trimmed body is the noise transfer function, i.e., the acoustic response at a selected position in the passenger compartment, e.g., driver's ear due to a mechanical excitation at a selected body mount. Detailed CAE models based on the FE method are today available for calculating this characteristic at low frequencies, corresponding to the engine idling and road excitation. However, the accuracy of CAE predictions of interior noise is still considered insufficient for the so-called analytical sign off, i.e., zero-prototypes vision. The current paper describes some investigations into the contribution of individual body panels to the overall interior noise. Also, the effect of selected interior trim items on the area investigated. Relative errors of prediction at different trim levels and different frequency ranges are discussed. Both experimental and CAE results are provided. The aim is to better understand the way the interior noise in vehicles is created, and how it can be controlled.

FRIDAY AFTERNOON, 1 DECEMBER 2006

HONOLULU ROOM, 3:45 TO 5:45 P.M.

Session 4pSAb

Structural Acoustics and Vibration: General Vibration and Measurement Technology

Peter C. Herdic, Cochair

Naval Research Lab., Physical Acoustics Branch, Code 7136, 4555 Overlook Ave., SW, Washington, DC 20375

Naoya Kojima, Cochair

Yamaguchi Univ., Dept. of Mechanical Engineering, 2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan

Contributed Papers

3:45

4pSAb1. Direct computation of degenerate elastodynamic solution of elastic wave propagation in a thick plate. Jamal Ghorieshi (Eng. Dept., Wilkes Univ., Wilkes-Barre, PA 18766)

The limiting form of elastodynamic solutions as frequency tends to zero leads to the elastostatic eigenvalue equations. However, this limiting procedure is not convenient. It is cumbersome when applied to the solutions obtained using Stokes potentials and, in the case of utilizing Lames potentials, it does not produce static solutions that are a function of position alone. In this paper it is shown that the exact solutions of elastostatic problems can, in general, be obtained in a straightforward manner by the use of harmonic potentials without recourse to any special limiting form of analysis. This method is applied to an infinite, elastic thick plate with traction-free parallel surfaces and the elastostatic eigenvalue equation. It is shown that the problem can be solved exactly in terms of harmonic functions, one of which is a scalar and the other one is a vector. It is noted that results are in agreement with the published solutions.

4:00

4pSAb2. Wave propagation characteristics of an infinite fluid-loaded periodic plate. Abhijit Sarkar and Venkata R. Sonti (Indian Inst. of Sci. Mech. Eng., IISc., Bangalore-560012, India)

A 1-D infinite periodic plate with simple supports placed along equidistant parallel lines is considered using the finite-element method. The plate is loaded with a finite-height fluid column covered on the top with a rigid plate. Results show a relation between the propagation constant of the fluid-loaded structure with its *in vacuo* counterpart. Since the acoustic medium is an additional wave carrier, the attenuation bands corresponding to the *in vacuo* structure turn out to be propagating. However, the presence of the fluid can also bring about attenuation regions within the *in vacuo* propagation bands. Primary propagation constants bring additional waves called space harmonics with them. Hence, a localized coincidence effect is seen where a particular harmonic falls below or above the acoustic wave number, leading to propagation or a mass loading effect. Occasionally, a complete attenuation band is created. This is verified by decomposing the

3288   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3288**

single span displacement profile into the space harmonics and also by computing the frequency response function (FRF) for a finite fluid-loaded periodic plate and observing the huge antiresonance dip in frequency in the exact same frequency band where an attenuation band was predicted for the infinite structure.

4:15

4pSAb3. The dynamic response of a plate subjected to various edge excitations. Baruch Karp (Faculty of Aerosp. Eng., Technion Israel Inst. of Technol., Haifa 32000, Israel)

Plane strain response of a semi-infinite, elastic plate to harmonic edge excitation is investigated analytically. The exact solution to this problem is obtained as a series expansion of the Rayleigh-Lamb modes of the plate. The variation of energy partition among the propagating modes with frequency of the edge excitation was found for load and displacement (symmetrical) perturbations of uniform and cosine form. The biorthogonality relation was employed in deriving the relative amplitudes of each mode to the given perturbation. The emphasis here is on the sensitivity of the far-field response, represented by propagating waves, to the details of the excitation. Within the frequency range investigated it was found that the plate's response is remarkably insensitive to whether the excitation is load or displacement type. The two types of edge excitation distributions, on the other hand, result in different patterns of energy partition above the first cutoff frequency with similar energy partition only within limited range of frequencies. The effect of the nature of the excitation on the dynamic response of the plate and a possible implication to dynamic equivalence is discussed.

4:30

4pSAb4. Measurement of structural intensity using patch near-field acoustical holography. Kenji Saijyou (Fifth Res. Ctr., TRDI, Japan Defense Agency, 3-13-1 Nagase, Yokosuka City, Kanagawa Prefecture, Japan, 239-0826)

Measurement of power flow in a structure, called the structural intensity (SI), is essential for vibration control and noise reduction. The near-field acoustical holography (NAH)-based measurement method is suitable to analyze the interrelationship between SI and acoustic intensity (AI) because NAH-based methods provide measurements of SI and AI simultaneously. Use of NAH requires the measurement of a pressure field over a complete surface located exterior to the structure. Therefore, if the measurement aperture is smaller than the structure, reconstructed results from the pressure on the finite aperture are seriously contaminated. This finite aperture effect prevents implementation of this NAH-based method on an actual large-scale structure such as a ship. Patch NAH and regularization method for SI measurement are applied to overcome this difficulty. This method enables implementation of the NAH-based SI measurement method in a large-scale structure. The effectiveness of this method is demonstrated by experiment.

4:45

4pSAb5. Flexural component and extensional component of vibration energy in shell structure. Taito Ogushi, Manabu Yahara, Masato Mikami, and Naoya Kojima (Dept. of Mech. Eng., Yamaguchi Univ., 2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan, j008ve@yamaguchi-u.ac.jp)

In this research, the behavior of the flexural component and the extensional component of vibration intensity and their transmission in curved shells are presented. L-shaped shell model was employed as an analysis model of FEM. As FEM analysis methods, both the frequency response analysis and the transitional response analysis were employed. The flexural component and the extensional component of vibration intensity (VI) were calculated by the results of FEM analysis. In the flexural component of the VI, the vibration energy supplied in the flat part decreased at the boundary from the flat part to the curved part and VI vectors flew in circumferential direction in the curved part. In the extensional component of the VI, the vibration energy appeared at the boundary from the flat part to the curved part and most VI vectors flew parallel to the shell axis in the curved part. The total vibration energy of the flexural component and the extensional component was conserved. So, the vibration energy transformed to each other between the flexural component and the extensional component in L-shaped shell.

5:00

4pSAb6. Seismic/acoustic detection of ground and air traffic for unattended ground sensor technology. Peter C. Herdic (Naval Res. Lab., Physical Acoust. Branch, Washington, DC 20375 and SFA Inc., Crofton, MD), Brian H. Houston, Phillip A. Frank, and Robert M. Baden (Naval Res. Lab., Washington, DC 20375)

Human footfall and vehicle traffic create surface waves in soil media that can easily be detected by seismic sensors. Field measurement data have been acquired with a triaxial geophone at several experimental sites. The in-plane-surface wave components dominate the response and decay at a rate of approximately $1/R$, where R is distance. This decay rate is due to the combined effect of spreading $(1/\sqrt{R})$ and damping losses in the soil. Further, the detection range is dependent upon the level of environmental noise, soil compliance, moisture content, and topography. Human detection was achieved in rural environments at distances up to ~30–40 m, and vehicle detection was possible at much greater distances. Seismic signals due to aircraft are small when compared to the acoustic signature. Ground-based microphone measurements clearly show the blade passage frequency tones of propeller airplanes and the broader band signature of turbojet aircraft. Time- and frequency-domain signal-processing methods for the detection and identification will also be introduced. These experimental results will be discussed with particular emphasis placed on wave phenomenon, detection and identification algorithms, and the related physics.

5:15

4pSAb7. Modeling of acoustic and elastic wave phenomena using plane wave basis functions. Tomi Huttunen, Jari P. Kaipio (Dept. of Phys., Univ. of Kuopio, P.O. Box 1627, FI-70211 Kuopio, Finland), and Peter Monk (Univ. of Delaware, Newark, DE 19716)

When simulating acoustic, elastodynamic, or coupled fluid-solid vibration problems using standard finite element (FE) techniques, several elements per wavelength are needed to obtain a tolerable accuracy for engineering purposes. At high wave numbers, the requirement of dense meshes may lead to an overwhelmingly large computational burden, which significantly limits the feasibility of FE methods for the modeling of wave phenomena. A promising technique for reducing the computational complexity is to use plane wave basis functions opposed to the low-order polynomials that are used in conventional FE methods. A possible method for utilizing the plane wave basis is the ultra-weak variational formulation (UWVF). The UWVF method can be used for acoustic Helmholtz problems, elastodynamic Navier problems, or fluid-solid systems characterized by a coupled Helmholtz-Navier model. A comparison of the UWVF technique with a low-order FE method shows reduced computational complexity and improved accuracy.

4p FRI. PM

Poster paper 4pSAb8 will be on display from 1:00 p.m. to 5:45 p.m. The author will be at the poster from 5:30 p.m. to 5:45 p.m.

4pSAb8. Application of vibration energy flow to evaluation of thickness. Akihiko Higashi (Dept. of Maritime Sci. and Technol., Japan Coast Guard Acad., 5-1 Wakaba-cho, Kure, Hiroshima, 737-8512, Japan)

In this study, the possibility of the useful application of the vibration energy flow is investigated. The vibration energy flow means the propagation of the vibration energy of the flexural waves. The vibration energy flow is expressed by the structural intensity. Here, it is easy to input the flexural waves in the thin plates and beam elements. Then, large structures such as ships use many of these thin plates and beam elements. But the usual methods of the evaluation and the inspection of these large structures are inefficient. Then, we investigated the possibility of the evaluation of the changed thickness of the structure by using the vibration energy flow analysis. As the result of analysis, the structural intensity suddenly changes at the position of the changed thickness. The changed quantity of the structural intensity corresponds to the change quantity of the thickness. Then, the evaluation method of the thickness of the structure is proposed. As a result, it was found that the change of the structural intensity indicates the change of the thickness. And the evaluation of the change of thickness of beams could be possible by using the proposed method.

FRIDAY AFTERNOON, 1 DECEMBER 2006

MOLOKAI ROOM, 1:00 TO 4:00 P.M.

Session 4pSC

Speech Communication: Variation in Production and Perception of Speech (Poster Session)

Heriberto Avelino, Cochair
Univ. of California—Berkeley, Dept. of Linguistics, 1203 Dwinelle Hall, Berkeley, CA 94720-2650

Haruo Kubozono, Cochair
Kobe Univ., Dept. of Linguistics, Faculty of Letters, Nada-ku, Kobe 657-8501, Japan

Contributed Papers

All posters will be on display from 1:00 p.m. to 4:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 2:30 p.m. and contributors of even-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

4pSC1. Cross-language perception of voice and affect. Christer Gobl, Irena Yanushevskaya, and Ailbhe N. Chasaide (Phonet. and Speech Lab., School of Linguistic, Speech and Commun. Sci., Trinity College Dublin, Dublin 2, Ireland, yanushei@tcd.ie)

The paper reports on a cross-language study of how voice quality and $f0$ combine in the signaling of affect. Speakers of Irish-English and Japanese participated in perception tests. The stimuli consisted of a short utterance where $f0$ and voice source parameters were varied using the LF-model implementation of the KLSyn88a formant synthesizer, and were of three types: (1) VQ only involving voice quality variations and a neutral $f0$ contour; (2) $f0$ only, with different affect-related $f0$ contours and modal voice; (3) VQ+$f0$ stimuli, where the voice qualities of (1) combine with specific $f0$ contours from (2). Overall, stimuli involving voice quality variation were consistently associated with affect. In (2) only stimuli with high $f0$ yielded high affective ratings. Striking differences emerge between the ratings obtained from the two language groups. The results show that not only were some affects consistently perceived by one language group and not the other, but also that specific voice qualities and pitch contours were associated with very different affects across the two groups. The results have important implications for expressive speech synthesis, indicating that language/culture-specific differences need to be considered. [This work is supported by the EU-funded Network of Excellence on Emotion, HUMAINE.]

4pSC2. An articulatory study of coronal consonants in Arrernte. Marija Tabain (LaTrobe Univ., Melbourne, Australia), Richard Beare (Monash Univ., Melbourne, Australia), Catherine Best (Univ. of Western Sydney, Sydney, Australia), and Louis Goldstein (Haskins Labs., CT)

This paper presents electro-magnetic articulography (EMA) data on the four coronal stops of Arrernte, an Australian language. The stops are: the lamino-dental "th," the apico-alveolar "t," the apico-postalveolar (or "retroflex") "rt," and the lamino-palatal "ty." Jaw, tongue tip (TT), and tongue body (TB) data were collected for two female speakers of the language. Results for the first speaker show a fronted tongue position for the laminal consonants, with the TT reflecting a similar location for both the dental and the palatal. However, for the palatal, the TB position is much higher, whereas for the dental, the TB is very low. For the apical consonants, the TT is not as far forward, and the TB is not quite as high as for the lamino-palatal. For both TT and TB, apico-postalveolar is further back than apico-alveolar. For the second speaker, the TT sensor failed, but in line with the first speaker, the TB sensor showed a higher position for the palatal. The other stops were lower and more forward, with the post-alveolar TB position higher than the laminal or alveolar stop position. For both speakers, the jaw position is lowest for the postalveolar. [Work supported by Australian Research Council and NIH: NIDCD.]

3290   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3290**

4pSC3. Symbolic phonetic features for pronunciation modeling.
Rebecca A. Bates,[a)] , Mari Ostendorf (Dept. of Elec. Eng., Univ. of Washington, Box 352500, Seattle, WA 98195), and Richard A. Wright (Univ. of Washington, Seattle, WA 98195)

A significant source of variation in spontaneous speech is due to intraspeaker pronunciation changes, often realized as small feature changes, e.g., nasalized vowels or affricated stops, rather than full phone transformations. Previous computational modeling of pronunciation variation has typically involved transformations from one phone to another, partly because most speech processing systems use phone-based units. Here, a phonetic-feature-based prediction model is presented where phones are represented by a vector of symbolic features that can be on, off, unspecified, or unused. Feature interaction is examined using different groupings of possibly dependent features, and a hierarchical grouping with conditional dependencies led to the best results. Feature-based models are shown to be more efficient than phone-based models, in the sense of requiring fewer parameters to predict variation while giving smaller distance and perplexity values when comparing predictions to the hand-labeled reference. A parsimonious model is better suited to incorporating new conditioning factors, and this work investigates high-level information sources, including both text (syntax, discourse) and prosody cues. Detailed results are under review with *Speech Communication*. [This research was supported in part by the NSF, Award No. IIS-9618926, an Intel Ph.D. Fellowship, and by a faculty improvement grant from Minnesota State University Mankato.] [a)]Currently at Minnesota State University, Mankato.

4pSC4. Acoustic phonetic variability and auditory word recognition by dyslexic and nondyslexic children.
Patricia Keating, Kuniko Nielsen (Phonet. Lab., Linguist., UCLA, Los Angeles, CA 90095-1543, keating@humnet.ucla.edu), Frank Manis, and Jennifer Bruno (USC, Los Angeles, CA 90089)

The hypothesis that dyslexia involves degraded phonological representations predicts impairments in behaviors that rely on these representations, such as auditory word recognition. Normal adult listeners recognize different pronunciations of a word as instances of the same lexical item, but more slowly and less accurately; dylexics should be even more impaired by acoustic phonetic variability. Children with and without dyslexia performed a word recognition task: on each trial, a child hears a target word, then eight probes (matching the target or not), responding yes/no to each probe. On some trials, probes are spoken by multiple talkers who differ in age, sex, speech style, etc.; on some trials the match probes also differ from the target in final stop consonant allophone. Responses are scored for accuracy and speed. Research questions include: Do all children demonstrate less accurate/slower recognition of words spoken by multiple talkers versus by one talker? Do all children demonstrate less accurate/slower recognition of words spoken with different allophones? Do dyslexic children demonstrate less accurate/slower recognition than nondyslexic children and, if so, for all trials, only for multiple talker trials, and/or only for different allophone trials; for all dyslexic children, or only those with particular phonological impairments? [Work supported by NIH.]

4pSC5. Intertalker differences in intelligibility of cochlear-implant simulated speech.
Tessa Bent, Adam B. Buchwald, and David B. Pisoni (Indiana Univ., Dept. of Psychol. and Brain Sci., 1101 E. 10th St., Bloomington, IN 47405, tbent@indiana.edu)

Are the acoustic-phonetic factors that promote highly intelligible speech invariant across different listener populations? Two approaches have been taken to investigate intelligibility variation for a variety of listener populations including hearing-impaired listeners, second language learners, and listeners with cochlear implants: studies on how speaking style affects intelligibility and other research on how inherent differences among talkers influence intelligibility. Taking the second approach, we compared intertalker differences in intelligibility for normal-hearing listeners under cochlear implant (CI) simulation ($n = 120$) and in quiet (n

$=200$). Stimuli consisted of 20 native English talkers' productions of 100 sentences. These recordings were processed to simulate listening with an eight-channel CI. Both clear and CI-processed tokens were presented to listeners in a sentence transcription task. Results showed that the most intelligible talkers in quiet were not the most intelligible talkers under CI simulation. Furthermore, listeners demonstrated perceptual learning with the CI-simulated speech but showed little learning in the quiet. Some of the acoustic-phonetic properties that were correlated with intelligibility differed between the CI-simulated speech and the speech in the quiet. These results suggest that the intertalker variations that result in highly intelligible speech observed in earlier studies are dependent on listener characteristics. [Work supported by NIDCD.]

4pSC6. The effect of phonological neighborhood density and word frequency on vowel production and perception in clear speech.
Rajka Smiljanic, Josh Viau, and Ann Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

Previous research showed that phonological neighborhood density and word frequency influence word recognition (Luce and Pisoni, 1998) and vowel production (Wright, 2002; Munson and Solomon, 2004; Munson, to appear), suggesting an interaction of lexical and phonetic factors in speech production and perception Here, we explore whether hyperarticulated, intelligibility-enhancing clear speech shows similar sensitivity to lexical-level structure. Nine American English talkers (five females, four males) produced 40 monosyllabic easy (frequent words with few lexical neighbors) and hard (infrequent words with many lexical neighbors) words in conversational and clear speech. Twenty-four subjects participated in a word-in-noise listening test. Results revealed a large effect of style on intelligibility and vowel production: words were more intelligible and vowels were longer and more dispersed in clear compared to conversational speech. Moreover, the female talkers produced larger vowel spaces than male talkers in both speaking styles. Vowels in hard words were marginally more dispersed than vowels in easy words in both speaking styles. However, within both speaking styles, easy and hard words were equally intelligible and of approximately equal duration. These results showed that phonetic properties of vowels were enhanced equally in clear speech regardless of their lexical properties.

4pSC7. Phoneme dependency of accuracy rates in familiar and unknown speaker identification.
Kanae Amino, Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan, amino-k@sophia.ac.jp), and Tsutomu Sugawara (Sophia Univ., Chiyoda-ku, Tokyo, 102-8554 Japan)

For perceptual speaker identification, the identification accuracy depends on the speech contents presented to subjects. Our previous studies have shown that stimuli containing nasals are effective for identifying familiar speakers [Amino *et al.*, Acoust. Sci. Tech. **27**(4) (2006)]. We have also presented the possibility that the interspeaker spectral distances reflect perceptual speaker similarities. In the present study, we conducted an experiment in which four unknown speakers were identified by 15 subjects. The stimuli were identical to those used in the previous study, in which ten speakers were identified by familiar listeners, although the speakers were fewer this time. Nine consonants in the CV structure were used as stimuli. The consonants were /d/, /t/, /z/, /s/, /r/, /j/, /m/, /n/, and /nj/; the vowel was restricted to /a/ for all CV syllables to simplify the experiment. The results showed that the nasals /n/ and /nj/ obtained higher scores. Tendencies in the differences among consonants were on the same order as those of the

previous experiment, but the average scores were lower than those for familiar listeners. [Work supported by Grant-in-Aid for JSPS Fellows 17-6901.]

4pSC8. Speech style and stereotypical character in Japanese. Akiko Nakagawa (Grad. School of Cultural Studies and Human Sci., Kobe Univ., 1-2-1 Tsurukabuto, Nada-ku, Kobe 657-8501, Japan, akiko.nakagawa@atr.jp) and Hiroko Sawada (Kyoto Univ., Kyoto 606-8501, Japan)

This study shows that ''stereotypical character'' is necessary to understand Japanese speech communication in addition to existing conceptions such as emotion, communicative strategy, register, and so on. Stereotypical character is here defined as a complex entity, consisting of information about gender, age, social status, physical features, characteristics, and speech style. The necessity of stereotypical character was shown through an auditory experiment involving a total of 70 speech sounds comprised of 15–19 short phrases (mean duration 1.4) selected from recordings of spontaneous speech of four adult female speakers of Japanese. Ten participants were asked to listen to these speech sounds randomly, and to classify them into four speakers. Each of the resulting auditory-perceptual categories was found to contain speech sounds from more than one speaker. Further analyses of these results suggested that the participants classified the speech sounds not according to invariant speaker characteristics but according to virtual stereotypical characters that are common in Japanese society. Therefore, such changeable speaker characteristics as ''busybody'' ''thoughtful,'' ''high-handed,'' and so on, can be elicited through speech sounds by Japanese speakers. [This work was partially supported by the Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (A), 16202006.]

4pSC9. Perceived vocal age and its acoustic correlates. Hiroshi Kido (Dept. of Commun. Eng., Tohoku Inst. of Technol., Taihaku-ku, Sendai, Japan 989-8577, kidoh@tohtech.ac.jp) and Hideki Kasuya (Intl. Univ. of Health and Welfare, Otawara, Japan 324-8501)

This study investigates relationships between perceived and chronological age of talkers and acoustic correlates of the perceived age. Most of the past studies were primarily concerned with the instability of the vocal-fold vibration extracted from sustained vowels. This study focuses on the dynamic nature of sentence utterances. Talkers included 115 healthy men, aged 20–60 years, who read a short sentence in Japanese. Listeners consisted of 70 men and women, aged 20–40 years, who made direct estimations of age. The results showed a strong correlation ($r = 0.66$) between the perceived and chronological age as well as the tendency toward overestimating the ages of younger talkers and underestimating those of older talkers, supporting past investigations [e.g., R. Huntley *et al.*, J. Voice **1**, 49–52 (1987)]. Acoustic parameters considered were median of the fundamental frequency ($F0$) contour, $F0$ range, declination of $F0$ contour, spectral tilt, median of the boundary frequencies above which irregularities dominate, and speaking rate. From both statistical graphical modeling and regression tree analysis, the speaking rate, $F0$ declination, and spectral tilt were found to be dominant acoustic correlates to the perceived age. [Work supported partly by a Grant-in-Aid for Scientific Research, JSPS (16300061).]

4pSC10. A cross-linguistic study of informational masking: English versus Chinese. Bruce A. Schneider, Liang Li, Meredyth Daneman (Dept. of Psych., Univ. of Toronto at Mississauga, Mississauga, ON, L5L 1C6 Canada, bschneid@utm.utoronto.ca), Xihong Wu, Zhigang Yang, Jing Chen, and Ying Huang (Peking Univ., Beijing, China 10087)

The amount of release from informational masking in monolingual English (Toronto, Canada), and Chinese (Beijing, China) listeners was measured using the paradigm developed by Freyman *et al.* [J. Acoust. Soc. Am. **106**, 3578–3588]. Specifically, psychometric functions relating percent-correct word recognition to signal-to-noise ratio were determined under two conditions: (1) masker and target perceived as originating from the same position in space; (2) masker and target perceived as originating from different locations. The amount of release from masking due to spatial separation was the same for English and Chinese listeners when the masker was speech-spectrum noise or cross linguistic (Chinese speech masking English target sentences for English listeners or English speech masking Chinese target sentences for Chinese listeners). However, there was a greater release from masking for same-language masking of English (English speech masking English target sentences) than for same-language masking of Chinese (Chinese speech masking Chinese target sentences). It will be argued that the differences in same-language masking between English and Chinese listeners reflect structural differences between English and Mandarin Chinese. [Work supported by China NSF and CIHR.]

4pSC11. Cross-linguistic differences in speech perception. Keith Johnson and Molly Babel (UC Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650)

This research explores language-specific perception of speech sounds. This paper discusses two experiments: experiment 1 is a speeded forced-choice AX discrimination task and experiment 2 is a similarity rating task. Experiment 1 was intended to investigate the basic auditory perception of the listeners. It was predicted that listeners' native languages would not influence responses in experiment 1. Experiment 2 asked subjects to rate the similarity between two tokens on a five-point equal interval scale; the purpose of this experiment was to explore listeners' subjective impression of speech sounds. In experiment 2 it was predicted that listeners' language would affect their responses. The same stimuli were used in both experiments. The stimuli consisted of vowel-fricative-vowel sequences produced by a trained phonetician. Six fricatives were used: /f, th, s, sh, x, h/. These fricatives were embedded in three vowel environments: /a_a/, /i_i/, and /u_u/. Tokens were presented to listeners over headphones with a 100-ms interval. Independent groups of 15 native Dutch and English listeners participated in each of the two experiments. Results suggest that listeners' language influenced responses in both experiments, albeit the result was larger in experiment 2. [Work supported by NIH.]

4pSC12. Neural coding of perceptual interference at the preattentive level. Yang Zhang (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, Minneapolis, MN 55455), Patricia Kuhl, Toshiaki Imada (Univ. of Washington, Seattle, WA 98195), Toshiaki Imada, and Masaki Kawakatsu (Tokyo Denki Univ., Inzai-shi, Chiba 270-1382, Japan)

Language acquisition involves neural commitment to language-specific auditory patterns, which may interfere with second language learning. This magnetoencephalography study tested whether perceptual interference could occur at the preattentive level. Auditory mismatch field (MMF) responses were recorded from ten American and ten Japanese adult subjects in the passive oddball paradigm. The subjects read self-chosen books and ignored the sounds. Three pairs of synthetic /ra-la/ syllables were used: one cross-category pair varied only in the third formant (F3), and the other two within-category pairs varied only in the second formant (F2). ANOVA results showed a main effect of acoustic dimension with significant interaction with subject groups ($p < 0.01$). As reported earlier, American listeners showed larger but later MMF responses for the F3 change. By contrast, Japanese listeners showed larger and earlier MMFs than Americans for changes in F2. Moreover, Japanese listeners had larger and earlier MMF responses for the changes in F2 as against changes in F3, which was more prominent in the right hemisphere than in the left. These results provided further support for the hypothesis that language experience produces neural networks dedicated to the statistical properties of incoming speech experienced in infancy, which later interfere with second language acquisition.

4pSC13. Russian and Spanish listener's perception of the English tense/lax vowel contrast: Contributions of native language allophony and individual experience. Maria V. Kondaurova (Program in Linguist., Purdue Univ., West Lafayette, IN 47907) and Alexander L. Francis (Purdue Univ., West Lafayette, IN 47906)

We examined the influence of listeners native phonology on the perception of American English tense and lax front unrounded vowels ([i] and [ɪ]). These vowels are distinguishable according to both spectral quality and duration. Nineteen Russian, 18 Spanish, and 16 American English listeners identified stimuli from a *beat-bit* continuum varying in nine spectral and nine duration steps. English listeners relied predominantly on spectral quality when identifying these vowels, but also showed some reliance on duration. Russian and Spanish speakers relied entirely on duration. Three additional tests examined listeners allophonic use of vowel duration in their native languages. Duration was found to be equally important for the perception of lexical stress for all three language groups. However, the use of duration as a cue to postvocalic consonant voicing differed due to phonotactic differences across the three languages. Group results suggest that non-native perception of the English tense/lax vowel contrast is governed by language-independent psychoacoustic factors and/or individual experience. Individual results show large variability within all three language groups, supporting the hypothesis that individual differences in perceptual sensitivity as well as the more frequently cited factors of second language education and experience play an important role in cross-language perception.

4pSC14. An analysis of acoustic deviation manner in spontaneous speech. Norimichi Hosogai, Kanae Okita, Takuya Aida, and Shigeki Okawa (Chiba Inst. of Technol., 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan)

Natural speech typically contains various phenomena deviated from the formal mode such as read speech. It is well known that those paralinguistic phenomena have an important role to give the human emotions and the state of the speakers in speech communication. This study attempts to extract the deviation as an acoustic ''vagueness,'' defined by temporal and dynamical acoustic features of speech. Especially the change of the vagueness during a certain period of speech, such as a 10-minute presentation, is focused. As the acoustic features, it used (*i*) modulation spectrum and (*ii*) syllable speed, which may have relations to the speech clarity and the tempo. For the experiments, 70 academic presentation speech data in the Corpus of Spontaneous Japanese (CSJ) are used. As the experimental results, significant properties in the patterns of the modulation spectrum and the syllable speed are obtained as a difference of the beginning and the ending periods of the presentation. This result will contribute to a human-like speech dialog system.

4pSC15. Nondurational cues for durational contrast in Japanese. Kaori Idemaru (Dept. of East Asian Lang. and Lit., Univ. of Oregon, 1248 Univ. of Oregon, Eugene, OR 97403) and Susan G Guion (Univ. of Oregon, 1290 Eugene, OR 97403)

This study explores potential secondary cues to a durational contrast by examining short and long stop consonants in Japanese. Durational contrasts are susceptible to considerable variability in temporal dimensions caused by changes in speaking rate. In this study, the proposal is examined that multiple acoustic features covary with the stop length distinction and that these features may aid in accessing the percept intended by the speaker, even when the primary cue, closure duration, is unreliable. The results support the proposal, revealing the presence of multiple acoustic features covarying with the short versus long contrast. Not only are there durational correlates to this contrast—the preceding vowel is longer and the following vowel is shorter for geminates than singletons—but there also are nondurational features covarying with this contrast. Greater fundamental frequency and intensity drops are found from the preceding to the following vowel for the geminate than the singleton stops. These results suggest the possibility that systematic variation of these acoustic

features is used in the perceptual categorization of the contrast in addition to the primary cue of closure duration. Moreover, the nondurational correlates are promising candidates for speech-rate resistant features.

4pSC16. Different motor strategies for increasing speaking rate: Data and modeling. Majid Zandipour, Joseph Perkell, Mark Tiede, Frank Guenther (M.I.T., Res. Lab Electron., Speech Commun. Group, 50 Vassar St, Cambridge, MA 02139, majidz@speech.mit.edu), Kiyoshi Honda (ATR Human Information Processing Res. Lab., Kyoto 619-0288, Japan), and Emi Murano (Univ. Maryland Dental School, Baltimore, MD, 21209)

Different motor strategies for increasing speaking rate: data and modeling EMG, kinematic and acoustic signals were recorded from two male subjects as they pronounced multiple repetitions of simple nonsense utterances. The resulting data indicate that the two subjects employed different motor strategies to increase speaking rate. When speaking faster, S1 significantly increased the size of the articulatory target region for his tongue movements, increased the speed of the tongue movements and the rate of EMG rise somewhat, while decreasing the movement duration significantly and movement distance slightly. In contrast, at the fast rate, S2 had the same size articulatory target region and rate of EMG rise as at the normal rate, but decreased the speed, distance, and duration of tongue movement slightly. Each subject had similar dispersions of acoustic targets in $F1-F2$ space at fast versus normal rates, but both shifted target centroids toward the center of the vowel space at the fast rate. Simulations with a biomechanical model of the vocal tract show how modulations of motor commands may account for such effects of speaking rate on EMG, kinematics, and acoustic outputs. [Work supported by NIDCD, NIH.]

4pSC17. Effect of speaking rate on individual talker differences in voice-onset-time. Rachel M. Theodore, Joanne L. Miller, and David DeSteno (Dept. of Psych., 125 NI, Northeastern Univ., 360 Huntington Ave., Boston, MA, 02115-5000, r.theodore@neu.edu)

Recent findings indicate that individual talkers systematically differ in phonetically relevant properties of speech. One such property is voice-onset-time (VOT) in word-initial voiceless stop consonants: at a given rate of speech, some talkers have longer VOTs than others. It is also known that for any given talker, VOT increases as speaking rate slows. We examined whether the pattern of individual differences in VOT holds across variation in rate. For example, if a given talker has relatively short VOTs at one rate, does that talker also have relatively short VOTs at a different rate? Numerous tokens of /ti/ were elicited from ten talkers across a range of rates using a magnitude-production procedure. VOT and syllable duration (a metric of speaking rate) were measured for each token. As expected, VOT increased as syllable duration increased (i.e., rate slowed) for each talker. However, the slopes as well as the intercepts of the functions relating VOT to syllable duration differed significantly across talkers. As a consequence, a talker with relatively short VOTs at one rate could have relatively long VOTs at another rate. Thus the pattern of individual talker differences in VOT is rate dependent. [Work supported by NIH/NIDCD.]

4pSC18. Variation in vowel production. Joseph Perkell, Majid Zandipour, Satrajit Ghosh, Lucie Menard (Speech Commun. Group, Res. Lab. of Electron., Rm. 36-511, M.I.T., Cambridge, MA 02139), Harlan Lane, Mark Tiede, and Frank Guenther (M.I.T., Cambridge, MA 02139)

Acoustic and articulatory recordings were made of vowel productions by young adult speakers of American English—ten females and ten males—to investigate effects of speaker and speaking condition on measures of contrast and dispersion. The vowels in the words teat, tit, tet, tat, tot, and toot were embedded in two-syllable ''compound words'' consisting of two CVC syllables, in which each of the two syllables comprised a real word, the consonants were /p/, /t/ or /k/, the two adjoining consonants were always the same, the first syllable was unstressed and the second stressed. Variations of phonetic context and stress were used to induce

dispersion around each vowel centroid. The compound words were embedded in a carrier phrase and were spoken in normal, clear, and fast conditions. Initial analyses of $F1$ and $F2$ on 15 speakers have shown significant effects of speaker, speaking condition (and also vowel, stress, and context) on vowel contrast, and dispersion around means. Generally, dispersions increased and contrasts diminished going from clear to normal to fast conditions. Results of additional analyses will be reported. [Work supported by NIDCD, NIH.]

4pSC19. Region, gender, and vowel quality: A word to the wise hearing scientist. Richard Wright (Dept. of Linguist., Univ. of Washington, Box 354340, Seattle, WA 98195-4340, rawright@u.washington.edu), Stephanie Bor, and Pamela Souza (Univ. of Washington, Seattle, WA 98105)

Sociophonetic research has established effects of regional accent and gender on spoken vowels. Many gender differences are due to sociolinguistic factors and thus vary by region. The implications for researchers and clinicians are important: gender variation must be controlled for according to the region of the listener and talker population. Moreover, speech perception stimuli used in research and in clinical applications have limited regional application. This poster illustrates these factors using the Pacific Northwest regional accent. The data, collected for a previous study on hearing aid processing, consist of three repetitions of eight vowels produced in real-word /h_d/ (or /_d/) contexts by six males and six females ranging in age from 19 to 60. Formants were measured using an LPC with an accompanying FFT and spectrogram for verification. The results revealed vowel-specific differences in the male and female speech over and above those typically associated with physiologic predictions, and different again from those observed in past studies from different regions. Taken as a whole, these data suggest that speech and hearing researchers should take care in selecting stimuli for general-use speech perception tests. [Work supported by NIDCD training grant (#DC00033) and NIH RO1 (1 RO1 DC006014).]

4pSC20. Acoustic characteristics of vowels in three regional dialects of American English. Ewa Jacewicz, Robert Allen Fox, Yolanda Holt (Speech Acoust. and Percept. Labs., Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210), and Joseph Salmons (Univ. of Wisconsin—Madison, Madison, WI)

Most of the comparative sociophonetic studies of regional dialect variation have focused on individual vowel differences across dialects as well as speaker variables. The present work seeks to define basic acoustic characteristics of entire vowel systems for three different regional variants of American English spoken in southeastern Wisconsin (affected by the Northern Cities Shift), western North Carolina (affected by the Southern Vowel Shift), and central Ohio (not considered to be affected currently by any vowel shift). Three groups of speakers (men and women) aged 20–29 years were recorded from each geographic area defined by two to three counties (creating a highly homogeneous set of speakers). Acoustic measures for the set of 14 monophthongs and diphthongs in /h_d/ context included vowel space area for each speaker, global spectral rate of change for diphthongized vowels (defined over the first three formant slopes), the amount of frequency change for F1 and F2 at two temporal points located close to vowel onset and offset (vector length), and vowel duration. These measures will establish both systemic and vowel inherent characteristics across the three dialects, serving as a basis for future examination of conditioning factors on vowels in chain shifts. Dialectal differences will be discussed. [Work supported by NIH NIDCD R01 DC006871.]

4pSC21. The rhythmic characterization of two varieties of Portuguese. Verna Stockmal, Emilia Alonso Marks, Audra Woods, and Z. S. Bond (Ohio Univ., Athens, OH 45701)

As spoken in Europe, Portuguese is said to be stress-timed, while Brazilian Portuguese appears to display characteristics of both stress and syllable timing [P. A. Barbosa, D.E.L.T.A. **16**, 369–402 (2000)]. We employed the Ramus *et al.* metric, based on acoustic-phonetic measurements [Ramus *et al.*, Cognition **73**, 265–292 (1999)], to investigate the possibility of distinguishing between the two varieties of the language. Five native speakers of European Portuguese and five native speakers of Brazilian Portuguese recorded the same short prose passage taken from a magazine. The talkers were asked to read at a normal, comfortable rate. The reading time of the passage averaged 60 s, with considerable differences among the talkers. From the vocalic and consonantal intervals, the Ramus metrics, percent vocalic interval and standard deviation of consonantal and vocalic interval, were calculated. The five talkers of the two language varieties differed on the values of these parameters. The values of %V and SD-V showed overlap between the two talker groups, while the BP talkers tended to show lower values for SD-C. Apparently, the rhythmic characterization of the two varieties of the language is not clearly categorical, but rather ranges along a continuum.

4pSC22. Indexical cues to talker sex and individual talker identity extracted from vowels produced in sentence-length utterances. Michael J. Berkowitz (Dept. of Psych., 301 Wilson Hall, Vanderbilt Univ., 111 21st Ave. South, Nashville, TN 37203, michael.j.berkowitz@vanderbilt.edu), Jo-Anne Bachorowski (Vanderbilt Univ., Nashville, TN 37203), and Michael J. Owren (Georgia State Univ., Atlanta, GA 30302)

The purpose of this study was to replicate and extend a previous study of indexical cuing [J.-A. Bachorowski and M. J. Owren, J. Acoust. Soc. Am. **106**, 1054–1063 (1999)] by including more vowel sounds spoken in more diverse contexts. Specific combinations of acoustic parameters that should represent talker sex and individual talker identity were identified using predictions based on known sources of variation in vocal production-related anatomy. This study utilized 100 recordings of sentence-length utterances, produced by each of 43 male and 44 female undergraduates, as well as 22 stock-phrase recordings produced by these same participants. One of five vowel sounds (/æ, ɛ, i, ə, u/) was isolated from each sentence and analyzed for F_0, $F1$, $F2$, $F3$, $F4$, vowel duration, jitter, shimmer, and harmonicity. Classification by talker sex was nearly perfect using a combination of cues related to both vocal-fold and vocal-tract anatomy. The accuracy of classification by individual identity depended strongly on cues relating to vocal tract-variation within sex.

4pSC23. Utterance-final position and projection of femininity in Japanese. Mie Hiramoto and Victoria Anderson (Dept. of Linguist., Univ. of Hawaii, 1890 East-West Rd., Honolulu, HI 96822)

Japanese female speakers frequent use of suprasegmental features such as higher pitch, longer duration, wider pitch range, and more instances of rising intonation *vis-a-vis* male speakers, is recognized as Japanese womens language (JWL) prosody. However, these features normally co-occur with gender-specific sentence-final particles (SFPs) like the strongly feminine "kashira." In this study, we examined the use of pitch and duration in utterances without SFPs, to determine whether JWL prosody is a function of SFPs or of utterance-final position. Eight male and eight female native Japanese speakers were instructed to read prepared sentences as though auditioning for a masculine theater role and then as though auditioning for a feminine role. Results indicate that utterance-final position is the projection point of JWL prosody even in the absence of SFPs. The data used for this study show high pitch, wide pitch range, long duration, and rising intonation at utterance-final positions when produced (by both men and women) in the feminine gender role. Conversely, in the masculine gender

role, both men and women avoid the use of such prosodic features, and may even avoid using rising intonation in interrogative sentences, where the tonal grammar calls for it.

4pSC24. Attitudinal correlate of final rise-fall intonation in Japanese. Toshiyuki Sadanobu (Kobe Univ., Tsurukabuto 1-2-1, Nada, Kobe, 657-8501, Japan)

Abrupt rise and subsequent fall intonation is common at the end of intonation units in Japanese, but its attitudinal correlate has not been fully elucidated yet. This intonation appears in the literature of the 1960's as politicians' way of speech, and nowadays not only politicians but many speakers including older generations often use it. However, this intonation is stigmatized as childish, and many people devalue it as an unintelligent way of speaking by young people. Where does this great gap between reality and image of this intonation come from? This presentation addresses this problem by focusing on natural conversation of Japanese daily life. The conclusions are as follows: (*i*) Rise-fall intonation often appears when the speaker talks about high-level knowledge, whereas it disappears when the speaker talks about their personal experience. (*ii*) Rise-fall intonation at the end of an intonation conveys the speaker's being so occupied with speaking that intonation unit. The childish image comes from the speaker's unawareness of their overall speech because of being occupied with local process. [Work supported by the Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (A), 16202006, and by the Ministry of Internal Affairs and Communications, SCOPE 041307003.]

4pSC25. Vowel devoicing in Japanese infant- and adult-directed speech. Laurel Fais, Janet Werker (Dept. of Psych., Univ. of BC, 2136 West Mall, Vancouver, BC V6T 1Z4 Canada, jwlab@psych.ubc.ca), Sachiyo Kajikawa, and Shigeaki Amano (NTT Commun. Sci. Labs., Seika-cho, Soraku-gun, Kyoto 619-0237 Japan)

It is well known that parents make systematic changes in the way they speak to infants; they use higher pitch overall, more pronounced pitch contours, more extreme point vowels, and simplified morphology and syntax (Andruski and Kuhl, 1996; Fernald *et al.*, 1989). Yet, they also preserve information crucial to the infants ability to acquire the phonology of the native language (e.g., phonemic length information, Werker *et al.*, 2006). The question examined in this paper is whether information other than phonemic segmental information is also preserved, namely, information concerning the phonological process of vowel devoicing. Devoicing of high vowels between voiceless consonants and word-finally after a voiceless consonant is a regular and well-attested phonological process in Japanese (Shibatani, 1990). A corpus of speech by Japanese mothers addressed to their infants and addressed to another adult was examined, and the degree and frequency with which they apply vowel devoicing in each type of speech was analyzed. Rates of vowel devoicing in speech to adults and infants are compared, accommodations made to infants and to hearing-impaired children are discussed (Imaizumi *et al.*, 1995), and the implications of these accommodations for the acquisition of vowel devoicing by Japanese infants are explored.

4pSC26. Language and gender differences in speech overlaps in conversation. Jiahong Yuan, Mark Liberman, and Christopher Cieri (Univ. of Pennsylvania, Philadelphia, PA 19104)

Language and gender differences in speech overlaps in conversation were investigated, using the LDC CallHome telephone speech corpora of six languages: Arabic, English, German, Japanese, Mandarin, and Spanish. To automatically obtain the speech overlaps between two sides in a conversation, each side was segmented into pause and speaking segments, and the overlap segments during which both sides were speaking were time stamped. Two types of speech overlaps are distinguished: (1) One side takes over the turn before the other side finishes (turn-taking type). (2)

One side speaks in the middle of the other side's turn (backchannel type). It was found that Japanese conversations have more short (less than 500 ms) turn-taking type of overlap segments than the other languages. The average number of such segments per 10 min of conversation for Japanese was 62.6 whereas the average numbers for the other languages ranged from 37.9 to 43.3. There were, however, no significant differences among the languages on the backchannel type of overlaps. Cross-linguistically, conversations between two females contained more speech overlaps (both types) than those between a male and a female or between two males, and there was no significant difference between the latter two.

4pSC27. An acoustic study of laryngeal contrast in three American Indian Languages. Heriberto Avelino (Dept. of Linguist., UC Berkeley, Berkeley, CA 94720-2650)

A contrast between modal and nonmodal phonation is commonly found in American Indian languages. The use of laryngealized voice has been reported in a number of languages from different linguistic families. This paper investigates the acoustics of laryngealized phonation in three indigenous languages spoken in Mexico, Yalalag Zapotec, Yucatec Maya, and Mixe. These languages differ in terms of the use of other features controlled by action of the larynx, i.e., tone. In Zapotec there is a contrast between high, low, and falling tones; Maya has phonemic high and low tones, whereas Mixe does not present phonemic pitch. The results show that the production of phonemic laryngeal vowels differs from language to language. The data suggest that the specific implementation of laryngealization depends in part on the relationship with contrastive tone. The patterns of the languages investigated provide new evidence of the possible synchronization of phonation throughout the vowel. With this evidence, a typology of modal/nonmodal phonation in phonation-synchronizing languages is proposed.

4pSC28. The comparison between Thai and Japanese temporal control characteristics using segmental duration models. Chatchawarn Hansakunbuntheung and Yoshinori Sagisaka (GITI, Waseda Univ., 29-7 Bldg. 1-3-10, Nishi-Waseda, Shinjuku-ku, Tokyo 169-0051, Japan, chatchawarnh@fuji.waseda.jp)

This paper compares the temporal control characteristics between Thai and Japanese read speech data using segmental duration models. The same and the different control characteristics have been observed from phone level to sentence level. The language-dependent and language-independent control factors have also been observed. In phone and neighboring phone level, different characteristics are found. Japanese vowel durations are mainly compensated by only adjacent preceding and following phones, which results from mora timing. Unlike Japanese, Thai vowel durations are affected by two succeding phones. It can be guessed that the differences come from syllabic structures. In word level, most content words tend to have longer phone durations while function words have shorter ones. In phrase level, both languages express duration lengtening of syllable/mora at the phrase initial and final. For language-specific factors, Thai tones express small alteration on phone duration. The comparisions explore the duration characteristics of the languages and give more understanding to be used in speech synthesis and second-language learning research. [Work supported in part by Waseda Univ. RISE research project of "Analysis and modeling of human mechanism in speech and language processing" and Grant-in-Aid for Scientific Research A-2, No. 16200016 of JSPS.]

4pSC29. Articulatory settings of French and English monolinguals and bilinguals. Ian L. Wilson (Univ. of Aizu, Tsuruga, Ikki-machi, Aizu-Wakamatsu City, Fukushima, 965-8580, Japan, wilson@u-aizu.ac.jp) and Bryan Gick (Univ. of BC, Vancouver, BC V6T1Z1 Canada)

This study investigated articulatory setting (AS), a language's underlying posture of the articulators. Interspeech posture (ISP) of the articulators (their position when motionless during interutterance pauses) was used as a measure of AS in Canadian English and Quebecois French.

4p FRI. PM

Optotrak and ultrasound imaging were used to test whether ISP is language specific in monolingual and bilingual speakers. Results show significant differences in ISP across the monolingual groups, with English exhibiting a higher tongue tip, more protruded upper and lower lips, and narrower horizontal lip aperture. Results also show that upper and lower lip protrusion are greater for the English ISP than for the French ISP, in all bilinguals who were perceived as native speakers of both languages, but in no other bilinguals. Tongue tip height results mirror those of the monolingual groups, for half of the bilinguals perceived as native speakers of both languages. Finally, results show that there is no unique bilingual-mode ISP, but instead one that is equivalent to the monolingual-mode ISP of a bilingual's currently most-used language. This research empirically confirms centuries of noninstrumental evidence for AS, and for bilinguals it suggests that differences between monolingual and bilingual modes do not hold at the phonetic level.

4pSC30. Temporal and spectral variability of vowels within and across languages with small vowel inventories: Russian, Japanese, and Spanish. Franzo F. LawII (Speech Acoust. and Percept. Lab., City Univ. of New York—Grad. Ctr., 365 Fifth Ave., New York, NY 10016-4309, flaw@gc.cuny.edu), Yana D. Gilichinskaya, Kikuyo Ito, Miwako Hisagi, Shari Berkowitz, Mieko N. Sperbeck, Marisa Monteleone, and Winifred Strange (City Univ. of New York—Grad. Ctr., New York, NY 10016-4309)

Variability of vowels in three languages with small vowel inventories (Russian, Japanese, and Spanish) was explored. Three male speakers of each language produced vowels in two-syllable nonsense words (VCa) in isolation and three-syllable nonsense words (gaC1VC2a) embedded within carrier sentences in three contexts: bilabial stops in normal rate sentences and alveolar stops in both normal and rapid rate sentences. Dependent variables were syllable duration and formant frequency at syllable midpoint. Results showed very little variation across consonant and rate conditions in formants for /i/ in Russian and Japanese. Japanese short /u, o, a/ showed fronting (F2 increases) in alveolar context, which was more pronounced in rapid sentences. Fronting of Japanese long vowels was less pronounced. Japanese long/short vowel ratios varied with speaking style (isolation versus sentences) and speaking rate. All Russian vowels except /i/ were fronted in alveolar context, but showed little change in either spectrum or duration with speaking rate. Spanish showed a strong effect of consonantal context: front vowels were backed in bilabial context and back vowels were fronted in alveolar context, also more pronounced in rapid sentences. Results will be compared to female productions of the same languages, as well as American English production patterns.

4pSC31. Does infant-directed speech in Tagalog resemble infant-directed speech in American English? Emmylou Garza-Prisby, Shiri Katz-Gershon, and Jean Andruski (Aud. & Speech-Lang. Pathol. Dept., Wayne State Univ., 207 Rackham, 60 Farnsworth St., Detroit, MI 48202)

This study compared the speech of a Filipino mother in infant- and adult-directed speech in order to investigate whether the mother used the acoustic features typically found in the infant-directed speech of American English-speaking mothers. Little acoustic documentation is available on the acoustic features of Tagalog, and no acoustic comparison of speech registers has so far been conducted. Impressionistically, Tagalog-speaking mothers' do not appear to use the features typically found in American mothers speech to their young infants. The mother was recorded talking to her infant and to another adult native speaker of Tagalog. Recordings were made in the mother's home and visits occurred during the first 6 months of the infant's life. Specific acoustic features analyzed include (a) vowel duration, (b) vowel format frequencies, (c) vowel triangle size, (d) rate of speech, (e) fundamental frequency, and (f) F0 range. Morphological and syntactic features were also analyzed, including (g) mean length of utterance and (h) sentence length. Results support a need for further study of speech registers in Filipino mothers.

4pSC32. Restricting relativized faithfulness and local conjunction in optimality theory. Haruka Fukazawa (Keio Univ., 4-1-1 Hiyoshi, Kohoku-ku, Yokohama, Japan)

Within the framework of optimality theory (OT), the two mechanisms, relativized faithfulness and local conjunction, play inevitable roles, especially when a simple constraint ranking fails to account for the data. However, their domain of application are too unrestricted and even overlapping each other. For instance, there are some cases which could be explained both by the ranking with relativized faithfulness and by the one with local conjunction. Syllable-final neutralization in German and geminate devoicing in Japanese loanwords are of interest in this context. The present paper proposes formal restrictions mostly on the local conjunction mechanism: the soundness of constraint combination, the number of constraints involved in a conjunction, and the local domain of conjunction. They not only can simplify the analysis but also give a more principled account for the overlapping cases. In fact, relativized faithfulness approach wins over local conjunction approach both in German neutralization and in Japanese loanwords. It is desirable for the universal grammar to be more restricted. Removing an overlap of theoretical devices is an important step toward the goal.

Session 4pUWa

Underwater Acoustics: Session in Honor of Leonid Brekhovskikh II

William A. Kuperman, Cochair

Scripps Inst. of Oceanography, Univ. of California, San Diego, Marine Physical Lab., La Jolla, CA 92093-0238

Oleg A. Godin, Cochair

NOAA, Earth System Research Lab., 325 Broadway, Boulder, CO 80305-3328

Invited Papers

1:00

4pUWa1. Underwater noise as source of information on conditions and dynamics of ocean environments. Alexander V. Furduev (N. N. Andreyev Acoust. Inst., 4 Shvernika St., Moscow 117036, Russia)

Leonid Brekhovskikh wrote in his book *The Ocean and the Human* (1987): ''Ocean noise is as important oceanographic parameter as temperature, current, and wind.'' Brekhovskikh created and headed the Laboratory of Acoustic Methods of Ocean Research in 1956. One of the scientific directions of the Laboratory was investigation of underwater noise both as interference for sound reception and a source of environmental information. Long-term studies on the unique acoustic research vessels created under the initiative of Brekhovskikh resulted in numerous important findings, including ambient noise spectra and envelopes of acoustic fluctuations, depth dependence of noise directivity, and mechanisms of ambient noise generation. Brekhovskikh was always eager to find practical applications of scientific results. Different methods of ensuring noise immunity of hydroacoustic arrays were developed under his supervision. Passive methods of acoustic navigation based on use of natural noise were suggested. Techniques for underwater acoustic monitoring of the ocean based either on ambient noise analysis or reemission of noise from a point away from the receiving system have been developed. The success of the team of scientists headed by Brekhovskikh was determined by the creative atmosphere around him: there was neither competition nor commercial interests. The common goal was knowledge of the ocean.

1:20

4pUWa2. Distributed acoustic sensing in shallow water. Henrik Schmidt (Ctr. for Ocean Eng., MIT, Cambridge, MA 02139)

The significance of Leonid Brekhovskikh to the ocean acoustics is undisputed. He was pioneering not only in terms of fundamental understanding of the ocean acoustic waveguide, but also the development of efficient and numerically stable approaches to propagation of sound in a stratified ocean. As such he has been an inspiration to a whole generation of model developers, leading to today's suite of extremely powerful wave theory models, capable of accurately representing the complexity of the shallow-water ocean waveguide physics. The availability of these computational tools have in turn led to major advances in adaptive, model-based signal-processing techniques. However, such computationally intensive approaches are not necessarily optimal for the next generation of acoustic sensing systems. Thus, ocean observation in general is currently experiencing a paradigm shift away from platform-centric sensing concepts toward distributed sensing systems, made possible by recent advances in underwater robotics. In addition to a fully autonomous capability, the latency and limited bandwidth of underwater communication make on-board processing essential for such systems to be operationally feasible. In addition, the reduced sensing capability of the smaller physical apertures may be compensated by using mobility and artificial intelligence to dynamically adapt the sonar configuration to the environment and the tactical situation, and by exploiting multiplatform collaborative sensing. The development of such integrated sensing and control concepts for detection, classification, and localization requires extensive use of artificial intelligence incorporating a fundamental understanding of the ocean acoustic waveguide. No other sources in literature provide this with the clarity and depth that is the trademark of Academician Brekovskikh's articles and classical textbooks. [Work supported by ONR.]

Contributed Papers

1:40

4pUWa3. When the shear modulus approaches zero: Fluids don't bend and Scholte leaves the room. Robert I. Odom (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698), Donna L. G. Sylvester, and Caitlin P. McHugh (Seattle Univ., Seattle, WA 98122-1090)

The 4×4 linear system of differential equations describing the propagation of the displacements and tractions in an elastic layered medium becomes singular as the shear modulus of the elastic medium approaches zero. There are a number of approximate ways to handle this singularity in order to impart numerical stability to the computation of the elastic waves in a layered medium. For example, one can impose an irrotational constraint on the displacements or introduce a massive elastic interface (MEI). Both of these ways of handling the weak shear strength are approximate, but avoid the need for singular perturbation theory [Gilbert, 1998]. Scholte waves are interface waves that propagate along the interface between an elastic solid and a fluid. They have nodes near or on the interface and decay exponentially into the bounding media. Scholte waves do not occur at the boundary between fluids. As the shear speed in the bounding elastic medium approaches zero, the Scholte waves disappear from the spectrum. We investigate this disappearance by applying singular perturbation theory

to the coupled fluid-elastic system. Among other things, we will address the rate in wave-number space at which the Scholte waves disappear from the spectrum.

<center>1:55</center>

4pUWa4. Measurement of the plane-wave reflection coefficient of the ocean bottom and the legacy of Leonid Brekhovskikh. George V. Frisk (Florida Atlantic Univ., 101 N. Beach Rd., Dania Beach, FL 33004) and Luiz L. Souza (Bose Corp., Framingham, MA 01701)

Leonid Brekhovskikh's classic text [*Waves in Layered Media* (Academic, New York, 1980)] inspired the development of several techniques for measuring the plane-wave reflection coefficient of the ocean bottom. Specifically, his application of the geometrical acoustics approximation to the problem of reflection of a spherical wave from a horizontally stratified medium provided the theoretical foundation for evaluating the strengths and weaknesses of various measurement methods. The most popular method assumes that the reflected field also consists of a spherical wave multiplied by the reflection coefficient evaluated at the specular angle. However, Brekhovskikh's work showed that this interpretation is confined to a limited range of angles and bottom structures and, if applied improperly, can lead to unphysical results such as negative bottom loss. This paper describes a technique which circumvents these deficiencies. It consists of measuring the pressure field magnitude and phase versus range due to a cw point source and Hankel transforming these data to obtain the depth-dependent Green's function versus horizontal wavenumber. The reflection coefficient is then obtained from the Green's function using the analytical relationship between the two quantities. The method is demonstrated using 220-Hz data obtained in a near-bottom geometry in the Icelandic Basin. [Work supported by ONR.]

<center>2:10</center>

4pUWa5. Field from a point source above a layered half-space; theory and observations on reflection from the seabed. Charles W. Holland (Appl. Res. Lab., The Penn State Univ., State College, PA)

L. M. Brekovskikh's book *Waves in Layered Media* has provided decades of graduate students and researchers alike with a comprehensive and enormously useful reference. One topic from that work, reflection from a point source above a plane-layered medium (the seabed), is discussed here. Both theoretical underpinnings and observations of reflection from a homogeneous halfspace, a transition layer with smoothly varying density and velocity profiles, and discrete layered media are considered for various shallow water sediment fabrics. [Work supported by the Office of Naval Research and NATO Undersea Research Centre.]

<center>2:25</center>

4pUWa6. Plane-wave model and experimental measurements of the directivity of a Fabry-Perot, polymer film, ultrasound sensor. Benjamin T. Cox and Paul C. Beard (Dept. of Med. Phys. and Bioengineering, Univ. College London, Gower St., London, WC1E 6BT, UK, bencox@mpb.ucl.ac.uk)

Optical detection of ultrasound is popular due to the small element sizes that can be achieved. One method exploits the thickness change of a Fabry-Perot (FP) interferometer caused by the passage of an acoustic wave to modulate a laser beam. This detection method can have greater sensitivity than piezoelectric detectors for sub-millimeter element sizes. The directivity of FP sensors and the smallest achievable effective element size are examined here. A plane-wave model of the frequency-dependent directional response of the sensor, based on Brekhovskikh's work on elastic waves in layered media, is described and validated against experimental directivity measurements made over a frequency range of 15 MHz and from normal incidence to 80 deg. In terms of applications, the model may be used to provide a noise-free response function that can be deconvolved from sound field measurements in order to improve accuracy in high-frequency metrology and imaging applications, or, for example, as a predictive tool to improve sensor design. Here, the smallest achievable effective element radius was investigated by comparing the directivity with that of a rigid circular pressure transducer, and found to be $\sim 0.9d$, where d is the thickness of the FP interferometer. [Funding was provided by the EPSRC, UK]

<center>2:40</center>

4pUWa7. The interference head wave and its parametric dependence. Jee Woong Choi and Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington)

The interference head wave that can exist in the presence of a sound-speed gradient in the sediment, is a precursor arrival representing a transition between the first-order head wave and the zeroth-order refracted wave. Using a parabolic equation (PE) simulation, Choi and Dahl [J. Acoust. Soc. Am. **119**, 2660–2668 (2006)] showed how the small shift in the dominant frequency of the interference head wave behaves as a function of the nondimensional parameter zeta, which itself is a function of center frequency, gradient, and range. For example, it was shown that the maximum frequency shift occurring in the vicinity of zeta equals 2. In this work, we investigate the amplitude and additional spectral properties of the interference head wave and analyze the cause of the frequency shift phenomenon using the ray theory. The limitation on the application of ray method also will be discussed. Finally, the conclusion will be verified by the time-dependent simulation using the RAM PE algorithm. [Work supported by the ONR.]

Session 4pUWb

Underwater Acoustics: Session in Honor of Fredrick Fisher

William A. Kuperman, Chair

Scripps Inst. of Oceanography, Univ. of California, San Diego, Marine Physical Lab., La Jolla, CA 92093-0238

Chair's Introduction—3:10

Invited Papers

3:15

4pUWb1. FLIP (Floating Instrument Platform): A major Fred Fisher contribution to ocean science. Fred N. Spiess, Robert Pinkel, William S. Hodgkiss, John A. Hildebrand (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD 0205, 9500 Gilman Drive; La Jolla, CA 92093-0205, fspiess@ucsd.edu), and Gerald L. D'Spain (UCSD, La Jolla, CA 92093-0205)

Frederick H. Fisher, a loyal and zealous member of the Acoustical Society of America, was an imaginative and effective developer of new techniques for research in both laboratory and seagoing acoustics. Most notable among his contributions was his work in bringing into being and enhancing the usefulness of the spar buoy laboratory, FLIP, from its inception in 1960. Not only did Fred use FLIP in his own research, its existence and many of its ancillary capabilities constituted a base for the seagoing research of others. The authors of this paper have benefited from FLIP's unique capabilities, starting with long-range sound propagation studies in the 1960's and 1970's. FLIP's stability and deep draft structure provided the platform for development of acoustic Doppler techniques for the measurement of ocean currents. Most recently, FLIP has been involved in studies of marine mammal vocalizations and use of multielement arrays to investigate details of shallow-water propagation. Fred's initial studies of sonar bearing accuracy, for which FLIP's construction was funded, and his dedication to advancing FLIP's ability to contribute to ocean science, constitute a legacy that is being utilized today, more than 40 years after FLIP's launching.

3:35

4pUWb2. Absorption of sound in seawater and ocean ambient noise, the scientific passions of Fred Fisher. John A. Hildebrand (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0205)

Fred Fisher made seminal contributions to ocean acoustics in the understanding of the absorption of sound in seawater and ocean ambient noise. Laboratory data and long-range sound propagation data revealed excess acoustic absorption in seawater. Fred Fisher spent much of his scientific career, beginning with his Ph.D dissertation, teasing out the contributions of various seawater components to sound absorption, and his work on this topic set the standard for understanding and modeling these phenomena. Ambient noise is an important aspect of underwater signal detection and is the focus of recent concerns about disturbance of marine organisms. Fred Fisher made important contributions to ambient noise studies by conducting measurements of vertical directionality, thereby testing models for ambient noise production. The value of archival ambient noise data and recent increases in ambient noise will be discussed.

Contributed Papers

3:55

4pUWb3. Fred Fisher's high-pressure work with eyewash and epsom salts. Christian de Moustier (Ctr. for Coastal & Ocean Mapping, Chase Ocean Eng. Lab, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824)

Starting in 1957 Fred Fisher led research programs devoted to high-pressure measurements related to the physical chemistry of sound absorption in seawater due to magnesium sulfate and other salts. As he put it, he spent his professional lifetime squeezing epsom salt. His interest in the low-frequency anomalous sound absorption in the ocean below 1 kHz led to the discovery of boric acid as the cause of the low-frequency relaxation. This paper is a short review of Fred Fisher's contributions to our knowledge of sound absorption in seawater, based in part on his carefully hand-written lecture notes and numerous low-pressure discussions.

4:10

4pUWb4. Fred Fisher and research with acoustic vector sensors; Marine Physical Laboratory's vertical array of directional acoustic receivers and ocean noise. Gerald L. D'Spain and William S. Hodgkiss (Marine Physical Lab, Scripps Inst. of Oceanogr., La Jollla, CA 93940-0701)

Fred Fisher had boundless enthusiasm for all topics acoustic. A chance encounter with him in the hallway usually led to a half-hour discussion of the latest research efforts at the lab and recent results he found exciting. In the 1980s, Fred became interested in the problem of identifying the physical phenomena forming the pedestal about the horizontal in vertical directionality measurements of the deep ocean's low-frequency noise field. Two competing mechanisms had been proposed: downslope conversion of coastal shipping and noise from high latitude winds coupling into the deep

4p FRI. PM

sound channel due to the shoaling of the sound channel axis. The relative contributions of these two mechanisms possibly could be separated if the azimuthal ambiguity of a vertical line array of hydrophones somehow could be broken. Therefore, Fred proposed to build a vertical array of "DIFAR" sensors, which led to the design and construction of the Marine Physical Lab's Vertical "DIFAR" Array. This talk will reminisce a bit about Fred as well as present some results from an ambient noise experiment conducted in 1992 on the continental shelf using the Vertical DIFAR Array co-deployed with MPL's freely drifting vector sensors, the Swallow floats. [Work supported by ONR and ONT.]

FRIDAY EVENING, 1 DECEMBER 2006   HAWAII BALLROOM, 7:00 TO 10:00 P.M.

Awards Ceremony

Anthony A. Atchley, President
Acoustical Society of America

Yôiti Suzuki, President
Acoustical Society of Japan

Acknowledgment of Honolulu Local Meeting Organizing Committees

Presentation of Fellowship Certificates

Anders Askenfeldt	James A. McAteer
Sergio Beristain	David R. Palmer
Philippe Blanc-Benon	Marehalli G. Prasad
David A. Conant	Hiroshi Riquimaroux
Andes C. Gade	Peter A. Rona
Anthony W. Gummer	Mark V. Trevorrow
Charles W. Holland	Michael Vörlander
Jody E. Kreiman	Joos Vos
Kevin D. LePage	Ben T. Zinn

Science Writing Award in Acoustics for Journalists to Radek Boschetty

Science Writing Award for Professionals in Acoustics to Edwin Thomas

Announcement of 2005 A. B. Wood Medal and Prize to Aaron Thode

Distinguished Service Citation to Thomas D. Rossing
Silver Medal in Noise to Alan H. Marsh
Silver Medal in Physical Acoustics to Henry E. Bass
Silver Medal in Psychological and Physiological Acoustics to William A. Yost
Wallace Clement Sabine Award to William J. Cavanaugh

Recognition of Acoustical Society of Japan meeting organizers
Recognition of Acoustical Society of America meeting organizers

ACOUSTICAL SOCIETY OF AMERICA
DISTINGUISHED SERVICE CITATION
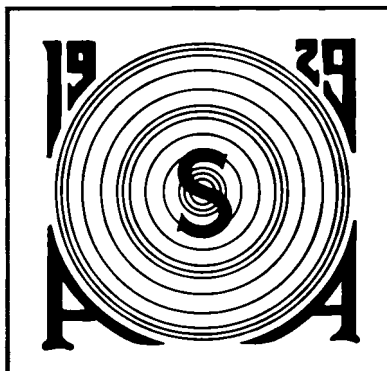
Thomas D. Rossing

2006

The Distinguished Service Citation is awarded to a present or former member of the Society in recognition of outstanding service to the Society.

PREVIOUS RECIPIENTS

Laurence Batchelder	1972	Frederick E. White	1987
Robert W. Young	1973	Daniel W. Martin	1989
Betty H. Goodfriend	1973	Murray Strasberg	1990
Gerald J. Franz	1974	William J. Cavanaugh	1994
Robert T. Beyer	1978	John C. Burgess	1996
Henning E. von Gierke	1978	Alice H. Suter	1997
R. Bruce Lindsay	1981	Elaine Moran	1999
William S. Cramer	1984	John V. Bouyoucos	2000
Stanley L. Ehrlich	1986	F. Avril Brenig	2000
Samuel F. Lybarger	1986		

CITATION FOR THOMAS D. ROSSING

...for contributions to the Society in bringing the joy of scientific discovery and knowledge of acoustics to young people, teachers, and Society members.

HONOLULU, HAWAII • 1 DECEMBER 2006

If you have attended one of his many presentations to high school students at an Acoustical Society meeting, you already know how well deserved this recognition is for Thomas D. Rossing. Students' faces light up every time Tom gets jump ropes to oscillate into their various modes or when he presents any one of a variety of fascinating sound demonstrations pulled from his ever-present suitcase. He truly has the rare ability to capture the interest of an audience of young students and to inspire their curiosity in exploring science. He has that unique skill to bring the joy of acoustics to young students. Likewise, if you have attended any of the special workshops where Tom has spoken, you also already know his dedication to the education of students and teachers. His talks and publications in acoustics over the years focus on this objective, aided by his skill in musical acoustics which provides an additional avenue to interest young students.

Most Acoustical Society members may not have had the pleasure of attending one of Tom's lectures or demonstrations but most members have read *ECHOES* since 1997 when he assumed the position of its editor. In this role he has contributed to increasing ASA members' general knowledge of acoustics through his tireless efforts to disseminate important information to all readers of *ECHOES*. A recent decision was made to publish *ECHOES* online so that it is more widely available and timely for readers. Over the past decade he has tracked down authors to write interesting articles and reported on ASA meetings and other related events in acoustics, including news about members covered in the ''We hear that'' section. Somehow he also finds time to edit the fall issue of the American Physical Society's *Forum on Education* newsletter!

Two other sections in *ECHOES*, ''Scanning the Journals'' and ''Acoustics in the News,'' provide readers with brief summaries of news and technical publications about acoustics that have appeared in a wide variety of publications which many readers would probably have missed reading otherwise. Compiling and writing these two sections of *ECHOES* requires the efforts of someone who likes to read and has a broad background and knowledge about acoustics and other fields of science. Tom possesses these qualities which he obtained through a career that has taken him around the world teaching, presenting papers, conducting research, and writing numerous articles and books.

As a mentor he has influenced scores of young scientists, and his laboratory has been a haven for numerous highly respected acousticians over the years. We do not have the space in this encomium to list all of his accomplishments, but in true ASA style, we refer the interested reader to the encomium written when Tom received the ASA's Silver Medal in Musical Acoustics in November 1992 (Vol. 92, No. 4, Pt. 2). Otherwise you may simply search ''Thomas D. Rossing'' on the web to get a sampling of his major publications.

Tom received a PhD in Physics in 1954 from Iowa State University, followed by 14 years as a member the faculty at St. Olaf College. Later he served as chair of Physics and Distinguished Research Professor at Northern Illinois University. He has been a visiting scientist at more than a dozen national and international universities—most recently at Seoul National University and the Korean Advanced Institute for Science and Technology. He is now a visiting professor at Stanford University. During his career Tom has written many tutorial papers on acoustics, several of which have appeared in *The Physics Teacher* published by the American Association of Physics Teachers (AAPT). Tom served as president of AAPT in 1991 and was awarded their Robert A. Millikan Award in 2000 which recognizes teachers ''who have made a notable and creative contribution to the teaching of physics.''

In 2003 the ASA's Rossing Prize in Acoustics Education was established to recognize outstanding educators in acoustics, funded through a generous donation from Tom. He has served on many ASA committees including Education in Acoustics, Standards, Public Relations, Medals and Awards, and the Technical Committee on Musical Acoustics (Chair 1984-1990). He has also presented two of ASA's tutorial lectures—the most recent one was a demonstration lecture in which he was assisted by ASA members and local high-school teachers.

He has organized many special sessions at Acoustical Society meetings on the subjects of musical acoustics and education in acoustics, including the ''take five'' sessions where presenters describe their favorite techniques and demonstrations for teaching acoustics. Tom also has organized special meetings, including the Catgut Acoustical Society meeting held in 1982 in DeKalb,

Illinois, the International Symposium on Musical Acoustics in Leavenworth, Washington in 1998 and a musical instrument workshop at the ASA's 2005 Spring meeting in Vancouver, Canada. Tom Rossing is the author or co-author of over 400 publications, including 15 books which include *The Science of Sound (3rd edition)*, *The Physics of Musical Instruments* (2nd edition, with Neville Fletcher), *Teaching Sound and Light,* and *Light Sciences Physics and the Visual Arts*. He is the editor for the *Handbook on Acoustics* which will be published in 2007 and co-author of the ASA's *Auditory Demonstrations* CD which is widely used for teaching the perception of sound.

Although Tom is well acquainted with so many fields of science and acoustics, his main love is musical acoustics. He has performed in barbershop quartets at ASA meetings, with his most recent performance at the 75th anniversary banquet in New York City in May 2004 when ASA met to celebrate the founding of the Acoustical Society of America in 1929. This was a good year for acoustics in that it was also the year of Tom Rossing's birth in Madison, South Dakota.

Whether Tom is singing on stage, striking a bell to show its modes, or immersed in some other acoustical demonstration, the smile on his face is a sure sign that he loves acoustics and wants to share that enthusiasm with all listeners. We know that the many readers of this encomium who know how much Tom has contributed to ''bringing the joy of scientific discovery and knowledge of acoustics to young people, teachers, and Society members,'' will share our joy in seeing him recognized by the Society for his many significant contributions to acoustics and to the Acoustical Society of America.

CHARLES E. SCHMID
UWE J. HANSEN

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Noise

Alan H. Marsh

2006

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Harvey H. Hubbard	1978	George C. Maling, Jr.	1992
Henning E. von Gierke	1981	Kenneth M. Eldred	1994
William W. Lang	1984	Larry H. Royster	1999
Tony F. W. Embleton	1986	Louis C. Sutherland	2002
William J. Galloway	1988		

CITATION FOR ALAN H. MARSH

. . . for contributions to the reduction of aircraft noise and for improvement to the quality of acoustical standards.

HONOLULU, HAWAII • 1 DECEMBER 2006

In the mid 1950's the major aircraft manufacturers in the United States were basing their futures on the development of commercial jet aircraft. Introduction of the Comet and Caravelle airplanes from Britain and France had brought a new and worrisome noise source into the communities of the world, causing the domestic airplane industry to begin to think for the first time about control of exterior noise as a major issue in their design considerations. In 1956 Alan Marsh joined Douglas Aircraft Company to begin a distinguished career in noise control engineering as one of the earliest pioneers in the control of aircraft noise.

Alan was born in Rochester, New York, on 21 March 1932, where he lived until enrolling at Williams College. After receiving a Bachelor of Arts degree in Physics from Williams he entered the Electrical Engineering program at the Massachusetts Institute of Technology (MIT). At MIT he came under the spell of enthusiasm for acoustics that Leo Beranek brought out in so many of us. Alan received a Bachelor of Science degree specializing in acoustics from MIT in 1956. Alan's penchant for learning continued after he moved to California and began his career at Douglas. Although he was working many hours of overtime Alan managed to pursue graduate work at the University of California at Los Angeles (UCLA) under the guidance of Alan Powell, then a professor of engineering at UCLA. Alan completed work and received a Master of Science degree in Engineering with a specialty in Aeroacoustics in 1962. His thesis involved the measurement of noise produced by a subsonic air jet impinging on a plane rigid surface.

At Douglas Alan was deeply involved in the earliest efforts to develop practical noise suppression techniques for turbojet engine exhaust noise for the DC-8 series airplanes. At about the same time, the development of the turbofan engine brought new noise control problems because of the strong tonal noise emitted by the turbofans and compressors of these engines. Development of sound-absorbing liners to reduce these noises became one of Alan's major projects. Noise control for these engines is greatly complicated by the high airflows to which materials are exposed, and the durability demands imposed by the airworthiness requirement for airplanes serving the public. Alan has confronted these competing design requirements throughout his career.

Development of reliable measurement instruments and procedures became an early interest for Alan. He became a strong and active participant in the development of standards for equipment and procedures at the national and international level through the standards program of the Acoustical Society of America (ASA), the Society of Automotive Engineers (SAE) Committee A-21 on Aircraft Noise, the International Standards Organization (ISO), and the International Electrotechnical Commission (IEC). He has been an individual expert on the ASA Committees S-1, Acoustics, and S-12, Noise, for many years. For ASA Alan has chaired or been a member of numerous standards working groups, including the standard for acoustical terminology and the development of editorial guidelines for the preparation of standards. For the IEC, Alan has been a U.S. member of the Working Group on Bandpass Filters, the Convener of the Working Group on Sound Level Meters (for the past twenty years), and other working groups, One of his most important contributions has been as the principal author of the SAE Aerospace Recommended Practice on the absorption of sound in air, which is now incorporated in U.S. and international standards for certification of aircraft for noise.

Alan has been a long-time contributor to the activities of the Acoustical Society since becoming a Member in 1956 and a Fellow in 1971. In addition to his service to the Society's standards program he has been Chair of the Technical Committee on Noise, a member of the Publication Policy Committee, an Associate Editor of the *Journal of the Acoustical Society of America*, a member of the Executive Council, and has served as technical chair for several of the semi-annual meetings of the Society. In addition to ASA, Alan was a founding member of the Institute of Noise Control Engineering of the USA for which he has served as President, Vice President, Secretary, Editor-in-Chief of the *Noise Control Engineering Journal*, technical chair and chair of INTER-NOISE Congresses. Alan has been, and currently is, an Officer, and a member of the Board of Directors of the International Institute of Noise Control Engineering.

After leaving Douglas in 1976, Alan formed DyTec Engineering to provide consulting engineering in aircraft and other forms of noise control. He has worked on numerous projects for a

variety of Douglas, Boeing, and other aircraft, as well as other noise control problems, and continues his active participation in acoustical standards development and editorial interests.

For his continuous contributions to the control of aircraft noise, his tireless participation in the development of acoustical standards, and his dedication to the improvement in the editorial quality of technical publications, Alan Marsh is a most deserving recipient of the Acoustical Society of America Silver Medal in Noise.

WILLIAM J. GALLOWAY

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Physical Acoustics

Henry E. Bass

2006

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Isadore Rudnick	1975	Allan D. Pierce	1991
Martin Greenspan	1977	Julian D. Maynard	1994
Herbert J. McSkimin	1979	Robert E. Apfel	1997
David T. Blackstock	1985	Gregory W. Swift	2000
Mack A. Breazeale	1988	Philip L. Marston	2003

CITATION FOR HENRY E. BASS

. . . for leadership in physical acoustics and contributions to the understanding of atmospheric sound propagation

HONOLULU, HAWAII • 1 DECEMBER 2006

Henry Ellis Bass is a native Sooner who more than typifies Oklahoma's motto, ''Labor conquers all things.'' Hank's success is cradled in his work ethic and in his insatiable curiosity and vision. His high school academic record earned him college scholarships even while working forty hours a week in support of his family. After his college studies were interrupted by a two-year stint in the Signal Corps, Hank returned to Oklahoma State University to complete his doctorate in physics with an ambitious investigation into rotational relaxation in polar polyatomic gases. Hank joined the Physics faculty of the University of Mississippi (UM) as an Assistant Professor in September 1970. Following promotions through the ranks, he was appointed a Frederick A. P. Barnard Distinguished Professor in 1988. In 1965, he joined the U.S. Army Reserve, serving as a staff officer in the Office of the Assistant Secretary of the Army for Research Development and Acquisition from 1971 until his retirement in 1993 as a Lieutenant Colonel. Hank formed the Physical Acoustics Research Group at the University of Mississippi. The group's reputation and research projects outgrew its facilities and Hank sought financial support for a new laboratory building. As a direct result of these efforts, the U.S. Congress established the National Center for Physical Acoustics (NCPA) in 1986. Hank was appointed Director in 1992.

Hank's research efforts are responsible for several significant findings and developments in physical acoustics. One of Hank's guiding principles is to bring talented people together to develop the best approach to solve a problem. Clear evidence of this approach is that Hank has shared authorship of his 119 peer-reviewed publications with ninety different colleagues. In his second year as Assistant Professor, he invited Hans Bauer to spend a year at UM. Hank recognized the wide applicability of Bauer's work and applied it to the problem of frequency and humidity dependence of sound absorption in the atmosphere. Hank's theoretical and computational models for atmospheric absorption of sound remain a benchmark in the field. His theories formed a major part of the background for the American National Standard, ''Method for Calculation of the Absorption of Sound by the Atmosphere.'' In 1978, Hank received the Acoustical Society's Biennial Award (now the R. Bruce Lindsay Award) in recognition ''for theoretical and experimental research on the effects of molecular relaxation and sound propagation in multicomponent gases.'' In subsequent years, he continued to extend, refine, and apply the fundamental knowledge gained in basic research. He applied the theoretical basis for the prediction of atmospheric absorption to the prediction of the propagation and development of the noise from sonic booms and explosion waves, as well as to the calculation of the acoustic energy production of lightning. Work on the acoustic energy spectrum of tornadoes then led to the development and marketing of the Tornado Alert System™, which identifies imminent tornadoes by their temporal and spectral characteristics.

Hank's study of atmospheric sound absorption led him to investigate ground effects on outdoor sound propagation. He carried out and led pioneering experiments that explored acoustic-to-seismic coupling, surface impedance and effects of ground cover. These experiments ultimately led to important findings on soil porosity, elasticity and near-surface layering including important developments such as long range detection of helicopters using the correlation between acoustic and seismic signals; improved methods for predicting noise propagation over a realistic ground surface; and an acoustic system for buried mine detection that was fully developed by James Sabatier, who, not surprisingly, was Hank's graduate student. In 1981, Hank organized the first International Symposium on Long Range Sound Propagation; particularly significant is the synergy between underwater and atmospheric acousticians that developed out of the symposia leading to new and improved theoretical, numerical and experimental techniques. Hank was also a founding member of NATO Research Study Group 11 that coordinates NATO research on passive acoustic detection.

Hank's directorship brought NCPA into the international arena where it gained a reputation for excellence in physical acoustics research. Because of this reputation, NCPA was chosen as lead institution for the installation of the U.S. Infrasound Monitoring Stations (IMS) that were established to support the United Nations' Comprehensive Test Ban Treaty. Hank assembled a team of infrasound experts from U.S. universities and national labs to oversee the installation of 11 stations, which are part of the global IMS network. Hank fostered the application of these arrays to

widely different phenomena such as earthquakes, volcano eruptions, meteors, and upper atmosphere physics. A nationally important application of the infrasound data was the investigation into the breakup of the space shuttle *Columbia*. Hank's contribution to the IMS carries import for future scientific applications of global-scale acoustics.

Speaking of the future, Hank has always declared, ''graduate students are our greatest product.'' Since 1994, Hank has been the director of the Physical Acoustics Summer School, which brings together graduate students, professionally mature scientists and distinguished lecturers to discuss topics students ordinarily would not encounter in the classroom.

A genuine individual, Hank sincerely appreciates talent and worth in others; his ''do what's best for the whole'' approach has made him a successful leader and administrator. For his students, employees and coworkers, he inspires loyalty and determination. Both things he readily gives in return. Hank is everybody's dad, brother, friend and mentor. As a good Oklahoman, he will help build your home, fix your well, or move your belongings to a new house. He also encourages people to seek their full, albeit hidden, potential. One of the authors of this encomium was an English major working as a secretary. Hank saw the potential and convinced her to switch to Physics. She is now pursuing a Master's.

Hank has an infectious laugh, playful wit and indefatigable curiosity. He motivates, stimulates, and inspires achievement in others. Sometimes, in the halls of NCPA, you can find him walking along with a jaunty gait and a twinkle in his eye all the while softly singing a Mac Davis song from the 80's, ''Oh Lord, its hard to be humble when you're perfect in every way.''

RICHARD RASPET
LIBBY FURR
ANTHONY ATCHLEY

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Psychological and Physiological Acoustics

William A. Yost

2006

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Lloyd A. Jeffress	1977	Nathaniel I. Durlach	1994
Ernest Glen Wever	1981	Neal F. Viemeister	2001
Eberhard Zwicker	1987	Brian C. J. Moore	2002
David M. Green	1990	H. Steven Colburn	2004
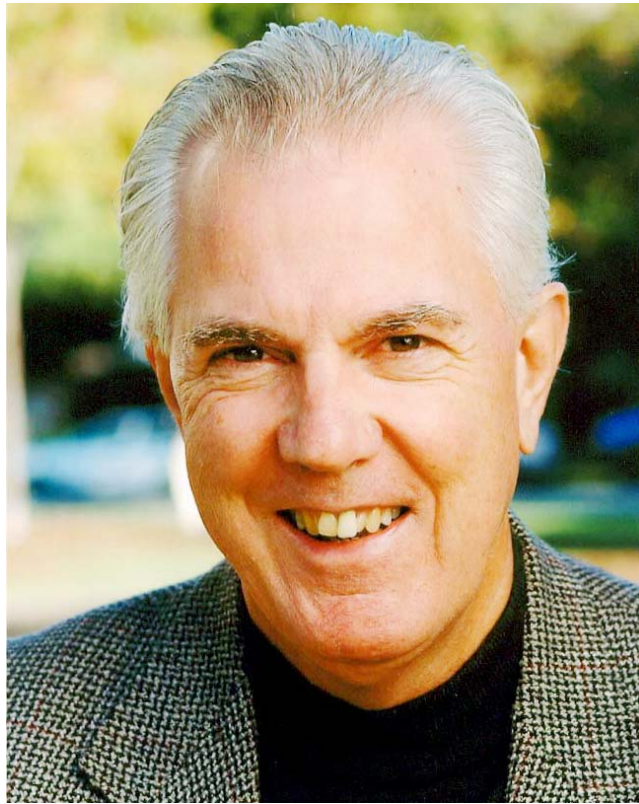
CITATION FOR WILLIAM A. YOST

. . . for contributions understanding pitch perception, sound source localization, and auditory processing of complex sounds.

HONOLULU, HAWAII • 1 DECEMBER 2006

Bill Yost has had a most distinguished career in acoustics, academics, and scientific research. He has made, and continues to make, substantial contributions to Psychological and Physiological Acoustics in the areas of pitch perception and pitch theory, binaural hearing and sound source localization, the detection and discrimination of amplitude modulation, and computational modeling of perception derived from temporal acoustic patterns. He has been a trailblazer in making all of us aware of the importance of sound source determination and segregation for the normal sense of hearing. At the same time, he has been an important person in the life of his home institution of 30 years, Loyola University Chicago, and has contributed significantly and gladly to the Acoustical Society of America, to other scientific and professional organizations and to national panels and government advisory groups. He is the most energetic academic scientist I know. I am proud to have him as a friend and colleague for the past 30 years, and to be able to write about his accomplishments.

Bill received a Ph.D. in Experimental Psychology from Indiana University in 1970, with an emphasis on the quantitative and mathematical study of hearing, in the tradition of behaviorism. Following a National Science Foundation post-doc with Dave Green at the University of California, San Diego, he spent seven years at the University of Florida in Gainesville. He came to Loyola University as Director of the Parmly Hearing Institute and Professor in the Department of Psychology where he thrived professionally and scientifically. Bill quickly transformed the Parmly Hearing Institute from, essentially, a laboratory and resource for the Director, into an internationally known research institute, with 25 faculty, students and staff, and with diverse approaches to hearing science. His research has been continuously funded since 1968 by grants from the National Institutes of Health, National Science Foundation, Air Force Office of Scientific Research, Office of Naval Research, Defense Advanced Research Projects Agency, and the Environmental Protection Agency. Under Bill's direction, the Parmly Hearing Institute received the largest research grant in Loyola University's history from the National Institute of Deafness and Other Communicative Disorders. He has published more than 300 scholarly articles, reports and book chapters. And while he was productive in the laboratory and in the classroom, Bill also served as the president of the ASA, as president of the Association for Research in Otolaryngology, as a Program Director for the National Science Foundation in Sensory Physiology and Perception, and later, as Loyola's Associate Vice President for Research and Dean of the Graduate School. We at The Parmly Hearing Institute are very pleased that now, Bill is ''just a faculty member'' in Psychology. I can think of no other active auditory researcher who has contributed to the diversity, quality and quantity as Bill Yost has contributed to modern psychoacoustic research. These research contributions have matched Bill's talent and energy as a scientific and academic administrator, spokesman, and educator for his field. There are three large themes of Bill Yost's research program: ripple noise processing, binaural signal detection and sound lateralization and localization, and envelope modulation processing including the phenomenon he and his colleagues recently established, known as 'modulation detection interference,' or MDI. There is no new symposium or review of psychoacoustics now that does not have to include or contend with Bill's seminal contributions to pitch perception, sound source localization, and MDI.

Of these several major scientific contributions Bill has made to his field, I focus on the analysis of pitch perception evoked by rippled noise. Rippled noise (sometimes also called repetition noise and cosine noise) is a signal having a sinusoidal amplitude spectrum that evokes a perception of pitch in human and other vertebrate listeners. This signal was originally used in hearing research, in part, because it provided an analog to the sinusoidal spatial frequency grating used in vision research. In effect, rippled noise produces a grating-like excitation pattern over the length of the cochlea, and was thought to permit an experimental investigation of the spatial pattern processing following cochlear frequency analysis. Bill began his experimental research program on rippled noise in 1978 by focusing on its pitch qualities, and the acoustic variables that contribute to pitch value and strength. After quantitatively investigating the effects of spectral modulation depth, component spacing, and harmonic versus inharmonic spectral ripples on pitch and pitch strength, Bill went on to quantitatively model these effects and then on to investigate the spectral dominance regions for evoking ripple noise pitch. More recently, Bill and his colleagues have studied

ripple noise perception neurophysiologically and psychophysically in other species. Most recently, work has focused on the question of time-domain versus frequency-domain processing in the generation of ripple noise pitch using iterated ripple noises. At the same time, Bill and his colleagues have developed an influential time-processing model of ripple noise pitch in which pitch strength is well predicted by the height of the first peak of the stimulus autocorrelation function. This body of work has been cumulative and systematic, resulting in some of the best evidence and theory regarding the temporal processing that may underlie pitch perception, echo processing, and spectral coloration in architectural acoustics. With characteristic energy and commitment, Bill has creatively and persistently exploited ripple noise signals to re-energize temporal pitch theory.

His honors and accomplishments are too numerous to mention in full, but include fellowship in the ASA, the American Speech-Language-Hearing Association, the American Psychological Society, and the American Association for the Advancement of Science. He was awarded an honorary Doctorate by his alma mater, The Colorado College, in 1997, and he serves as a National Associate of the National Academy of Sciences—National Research Council.

It is remarkable that for such an important leader in his field, Bill has also had a rich personal and family life. He and his wife Lee are world travelers, and Bill is a talented athlete on the tennis court, the fairways, and the ski slopes. Bill and Lee have raised two lovely and successful daughters, Kelley and Alyson, and now are enthusiastic new grandparents of Zachary Michael Abrams. Congratulations, Bill, for this richly deserved Silver Medal in Psychological and Physiological Acoustics.

RICHARD R. FAY

WALLACE CLEMENT SABINE AWARD
OF THE
ACOUSTICAL SOCIETY OF AMERICA

William J. Cavanaugh

2006

The Wallace Clement Sabine Award is presented to an individual of any nationality who has furthered the knowledge of architectural acoustics, as evidenced by contributions to professional journals and periodicals or by other accomplishments in the field of architectural acoustics.

PREVIOUS RECIPIENTS

Vern O. Knudsen	1957	Cyril M. Harris	1979
Floyd R. Watson	1959	Thomas D. Northwood	1982
Leo L. Beranek	1961	Richard V. Waterhouse	1990
Erwin Meyer	1964	A. Harold Marshall	1995
Hale J. Sabine	1968	Russell Johnson	1997
Lothar W. Cremer	1974	Alfred C. C. Warnock	2002

SILVER MEDAL IN
ARCHITECTURAL ACOUSTICS

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENT

Theodore J. Schultz   1976
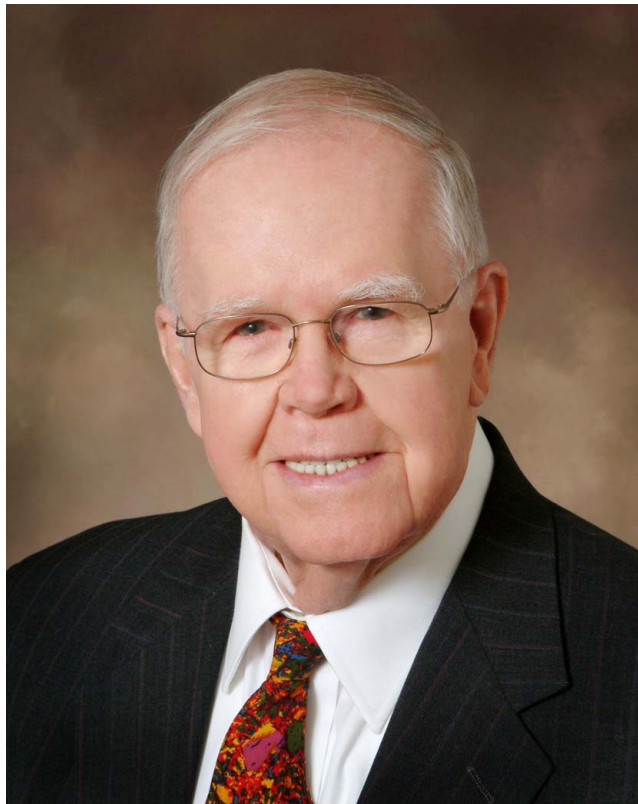
CITATION FOR WILLIAM JOSEPH CAVANAUGH

. . . for contributions to the practical application of architectural acoustics in building design and to education in architectural acoustics

HONOLULU, HAWAII • 1 DECEMBER 2006

Bill has changed the world as we hear it. Although always eager to redirect credit to his colleagues, most of us know at least some of the roles he played that have helped to define the entire field of architectural acoustics.

William J. Cavanaugh and the Acoustical Society of America (ASA) were both born in 1929– Bill in Boston. He can still recite much of the history of the Bunker Hill Monument that at 8 and 9 years old he would hawk to tourists, and sometimes with a touch of exaggerated ''Bostonese'' that often earned an extra nickel or two. As a growing boy, he juggled various jobs such as delivering newspapers, tomato-packaging at the old Quincy Market; the Depression was ''just fine because no one knew they were poor.'' During World War II he attended The English High School, Boston's public ''prep'' school for students who had dreams of attending nearby MIT, and, as the second oldest of six brothers and a sister grew to be the tallest.

In 1951 Bill earned his Architecture Degree from the Massachusetts Institute of Technology (MIT), and was awarded the American Institute of Architects Student Medal. With four years of Reserve Officer Training Corps (ROTC) at MIT, Bill was immediately ordered to active duty with the U.S. Army Corps of Engineers serving until the Korean War armistice in 1953. He remained in the Army Reserves until 1982, retiring with the rank of Colonel.

Returning from service and recently married to his life-mate, Ginny, he searched for a job in architecture. Settling for a job at Polaroid in their Plant Engineering Department, his MIT class-mate Ranger Farrell suggested that Bill would thrive at the rapidly growing Cambridge office of Bolt Beranek and Newman (BBN) which was established in 1948 to consult on the acoustics of the United Nations Headquarters complex in New York City. Bill joined BBN in February of 1954.

BBN's founding partners, Dick Bolt, Leo Beranek, and Bob Newman as well as many of the extraordinarily talented BBN staff, would have powerfully-positive impacts on Bill's acoustics career and on the entire field of architectural acoustics consulting. Dick Bolt as Professor of Physics and Director of the Acoustics Laboratory was central to all things acoustical, especially around MIT. Leo Beranek stimulated his students at MIT as well as everyone on the BBN staff to publish results of projects and supporting research. Bob Newman, a remarkable and animated teacher, became Bill's immediate mentor, and encouraged Bill's own teaching, always emphasiz-ing expressive language that would be persuasive to architects and engineers, and always with an irrepressible enthusiasm. Bill credits this engaging teaching style as the principal reason that BBN was so successful in architectural acoustics consulting, proving that educated clients are the best clients.

Bill soon began teaching acoustics classes at the MIT School of Architecture, the Boston Architectural Center, the Rhode Island School of Design, and more. Many students became future clients, and some even became acousticians. His passion for ''spreading the word'' eventually, following the untimely death of Bob Newman in 1983, led to the establishment of the Robert Bradford Newman Student Award Fund which sponsors Newman Student Medal Awards for excellence in the study of architectural acoustics and the Theodore J. Schultz Grants for excellence in teaching, both overseen by the Technical Committee on Architectural Acoustics (TCAA) of the Society.

Acoustics books and teaching aids have benefited from Bill's influence as well, in spite of localized pockets of resistance. In the late 1970's, Bill proposed to the Society that the poster sessions on various building types at ASA meetings could be a rich resource for Society publica-tions. Doubts were expressed, but Bill's enthusiasm and persistence were unstoppable. In 1982, ASA architectural acoustics session posters were collected into the first paperback ''Halls for Music Performance: Two Decades of Experience, 1962-1982,'' edited by Rick Talaske, Red Wetherill, and Bill Cavanaugh. The fifth in this ''poster book'' series, ''Halls for Music Perfor-mance: Another Two Decades of Experience, 1982-2002'' was produced in 2003—an elegant hard-cover, color-printed edition. Two more posterbooks in the series are in process. Bill's repu-tation has generated many invitations to write chapters, forewords and articles on architectural acoustics and noise control directed to the building professions following in the footsteps of Wallace Clement Sabine in whose honor the medal that Bill will be receiving today is named.

Immersed in a rapidly growing, diverse and intense environment, Bill's tenure at BBN was the

defining period of his career. His willingness to assist and counsel others at all levels of technical and administrative tasks became a trademark of his leadership style. For example, not long after Bill joined the BBN consulting staff he inherited the unenviable task of collecting timesheets and getting bills out to clients for the architectural acoustics group. Less than ten years later Bill was named Divisional Vice President and Director of BBN's Architectural Technologies Division.

Along with teaching and management matters, he maintained high chargeability on consulting projects and inevitably found an interesting ''wrinkle'' in almost every project. One particularly-heretical wrinkle was the concept that adding background noise could improve acoustics; after all, ''silence is golden.'' However, Bill and his BBN colleagues repeatedly demonstrated the benefits of introducing properly-tailored background sound particularly in the ever growing number of open-plan offices being constructed thoughout the country. Research gleaned from consulting projects as well as from laboratory experiments prompted optimizing the spectra and level of background sound, quantifying other speech-privacy parameters, and developing simplified methods for design and analysis of acoustical privacy problems and, more importantly, explaining the counter-intuitive concept of ''masking'' to clients. The resulting seminal publication was ''Speech Privacy in Buildings,'' by W.J. Cavanaugh, W.R. Farrell, P.W. Hirtle, and B.G. Watters (J. Acoust. Soc. Am. 34, pp. 475-492 (1962)). Many others followed. Speech privacy remains a high priority, especially with new federal mandates for privacy in health-care facilities; the most recent ASA meeting (spring 2006 in Providence, RI) featured special sessions, meetings, and a dinner banquet to celebrate ''Fifty Years of Speech Privacy.''

Most acoustical consulting firms can trace a relationship to BBN, either through alumni or outgrowth firms, which often means a connection to Bill Cavanaugh. Remaining close to his many colleagues, Bill left BBN in 1970 to establish an independent consulting practice. This led almost immediately to a partnership with Lawrence Copley, Cavanaugh Copley Associates (CCA), to handle large consulting assignments particularly acoustical environmental impact studies.

In 1975, Bill and Lawrie mututually decided to terminate the partnership and Bill partnered with Greg Tocci, CCA's only full time employee, to found Cavanaugh Tocci Associates (CTA) which today is a thriving office of 16 professional staff members in Sudbury, MA undertaking consulting and research assignments in architectural, structural, mechanical and environmental acoustics. Throughout the variety of projects ''in and around and away-from buildings,'' Bill's wide experience, his wise counsel, and his overall influence is simply overwhelming.

In addition to speech privacy, Bill is particularly prominent in developing acoustical criteria for cinemas and for outdoor concert venues. In the early 1980's, owners of rapidly growing movie theater chains were vitally interested in improving the acoustical experience for audiences; ensuing evaluations of cinemas across the country led to criteria that are meaningful to most architects and engineers, and are now the foundation for quantifying appropriate sound isolation, HVAC noise, and theater surface treatments. Bill's interest in the thorny problems of neighbor complaints of sound from outdoor amphitheaters led to and continues to influence the entire field of outdoor concert facility design, as well as methods to improve relations with affected communities.

Bill's service to the ASA includes Member of the Executive Council, Chair of the Technical Committee on Architectural Acoustics, and chair or member of administrative committees including Regional Chapters, College of Fellows, Public Relations, Archives and History, and Rules and Governance. He was awarded the ASA Distinguished Service Citation in 1994. Bill served as president of the National Council of Acoustical Consultants (NCAC 1977-79), and was awarded their C. Paul Boner Medal in 1983 for ''Distinguished Contributions to the Acoustical Consulting Profession.'' He also served as president of the Institute of Noise Control Engineering (1993). His work with local chapters and other organizations is legendary, and he has earned a reputation as a skilled negotiator, if not peacemaker, in difficult disputes. At any meeting, Bill is among the best people to locate for almost any kind of assistance.

Ginny, his five children (including Bill, Jr. who died tragically in 1982 while on active duty in the US Air Force), and a growing number of grandchildren are the love of his life. He is also in love with acoustical consulting, and it shows.

His countless friends and admirers, the Technical Committee on Architectural Acoustics, and the entire Acoustical Society of America are proud to congratulate Bill on this richly-deserved Wallace Clement Sabine Medal.

Bill Cavanaugh, CONGRATULATIONS and THANK YOU!

K. ANTHONY HOOVER
LEO L. BERANEK

Session 5aAA

Architectural Acoustics, Speech Communication, and Psychological and Physiological Acoustics: Psychological Aspects of Speech in Rooms I

John S. Bradley, Cochair
National Research Council, Inst. for Research in Construction, Ottawa, Ontario K1A 0R6, Canada

Hiroshi Sato, Cochair
National Inst. for Advanced Industrial Science and Technology, Inst. for Human Science and Biomedical Engineering, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566 Japan

Chair's Introduction—7:30

Invited Papers

7:35

5aAA1. Why we should design for optimum reverberation times in rooms for speech communication. J. S. Bradley (Inst. for Res. in Construction, Natl. Res. Council, Montreal Rd., Ottawa, Canada K1A 0R6)

Various texts list recommendations for optimum reverberation times as ideal goals in rooms for speech communication. However, some newer documents, such as ANSI S12.60, talk about maximum recommended reverberation times in rooms for speech communication. The source of the changed approach can be traced to interpretations of experimental results that do not consider the full implications for speech communication in real rooms, and also to the fact that reverberation time is not an ideal predictor of the quality of speech communication. In the extreme, minimizing reverberation times would lead to near anechoic rooms for speech and inadequate signal-to-noise ratios. The need for optimum reverberation times can be explained as a simple need to first achieve adequate signal-to-noise ratios. However, a more complete understanding is obtained by examining the benefits of early-arriving reflections of speech sounds on the intelligibility of the speech to listeners. Attempts to determine optimum reverberation times for normal hearing listeners, which are based on a balance between avoiding excessive reverberation and maintaining adequate signal-to-noise ratios, lead to a range of acceptable values that can vary with the ambient noise level. This discussion will be supported with examples from room acoustics measurements in classrooms and meeting rooms.

7:55

5aAA2. The relation between speech transmission index, clarity, and reverberation time and listening difficulty in the impulse response database of AIJ. Hiroshi Sato (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan), Yoshio Nishikawa (Konoike Co., Ltd., Tsukuba 305-0003, Japan), Hayato Sato, and Masayuki Morimoto (Kobe Univ., Nada, Kobe 657-8501, Japan)

The Speech Communication Research Working Group of AIJ (Architectural Institute of Japan) is collecting information on rooms including digitized impulse responses (IRs) to establish a database for evaluating and designing the speech transmission quality of rooms. This database consists of 966 measured IRs. This study presents the relationships between speech transmission index (STI), clarity (Cx) and reverberation time (T) as measures to consider for the design and evaluation of the speech transmission performance of rooms. The data show a wide range of STI and Cx values for a given T, and the minimum STI at each reverberation time can be obtained by diffused field theory. STI and Cx are seen to be better indicators than T for the design of rooms for speech and that T is not as good. Relationships between STI, Cx, and listening difficulty ratings from previous studies [Proc. of RADS (2004), Proc. of Forum Acusticum, pp. 1713–1718 (2005)] found both STI and Cx can be used as predictors of listening difficulty ratings. Finally, listening difficulty ratings of individual IR in the database are estimated and the distribution of listening difficulty ratings in a variety of rooms are presented as a solid bases to design speech transmission quality of rooms.

8:15

5aAA3. Unifying approaches for modeling and predicting speech intelligibility. Adelbert W. Bronkhorst and Sander J. van Wijngaarden (TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands, adelbert.bronkhorst@tno.nl)

Speech intelligibility is practically always affected by the acoustic environment (reverberation, interfering speech, noise). When speech is processed electronically, distortions introduced by the transmission chain can also deteriorate intelligibility. Furthermore, intelligibility can be strongly influenced by nonacoustic factors such as contextual information, non-nativeness, and hearing impairment. Various intelligibility prediction methods have been developed (including SII, STI, SRS, and PESQ), which cover almost all factors influencing intelligibility. However, each individual method has limited applicability and no attempts have been made to unify them. Recently, the application domains of two widely used methods, the SII and the STI, have been significantly extended. They now, for example, cover interference that fluctuates in level, live speech, and digital transmission channels. These methods are actually very

5a SAT. AM

similar in the way in which they treat frequency-domain effects. Because of this, it is not difficult to fuse the STI and SII into a single method with an unparalleled application domain. Still, important factors such as informational masking and context effects cannot be addressed and more effort should be invested in combining the SII/STI with other, less similar methods (such as the SRS) in order to develop a single, powerful, standard model for predicting speech intelligibility.

8:35

5aAA4. Influence of word familiarity on spoken word recognition. Shigeaki Amano, Tadahisa Kondo (NTT Commun. Sci. Labs., NTT Corp., 2-4 Hikari-dai, Seika-cho, Souraku-gun, Kyoto 6190237, Japan, amano@cslab.kecl.ntt.co.jp), Shuichi Sakamoto, and Yôiti Suzuki (Tohoku Univ., Japan)

Spoken word recognition is affected by many factors including sound pressure level, signal-to-noise ratio, word familiarity, word frequency, and phoneme sequence plausibility. Of these factors, word familiarity is assumed to have a strong effect but has received little attention. One reason for this is that word familiarity data for most words have been unavailable. However, we can now confirm the word familiarity effect because in 1999 Amano and Kondo published a word familiarity database for about 80 000 Japanese words. Experimental results clearly show that spoken words with a high familiarity are recognized better and faster than those with a low familiarity. In addition, word familiarity correlates with the recognition score of spoken words much better than word frequency, which was previously regarded as an influential factor in word recognition. Because the word familiarity effect is so strong, it is very important to control the familiarity of spoken words to obtain stable and reasonable results when assessing the speech hearing ability of individuals, and when measuring speech clarity in communication media and public spaces.

8:55

5aAA5. Development of word lists for word intelligibility tests in Japanese. Shuichi Sakamoto, Yôiti Suzuki (Res. Inst. of Elec. Commun., Grad. School of Info. Sci., Tohoku Univ., 2-1-1, Katahira, Aoba-ku, Sendai, 9808577, Japan, saka@ais.riec.tohoku.ac.jp), Shigeaki Amano, and Tadahisa Kondo (NTT Corp., Seika-cho, Soraku-gun, Kyoto, 6190237, Japan)

Speech intelligibility tests are widely used to evaluate personal speech hearing ability accurately. Although various lists have been proposed for the tests, almost all lists in Japan do not properly consider word difficulty as an aspect of speech recognition. Consequently, even if these tests are performed under identical conditions, correct results cannot be obtained. We have been developing word lists for word intelligibility tests to cope with these problems. We control word familiarity, which strongly affects word recognition scores, to equalize difficulty in recognition between word lists. Words are divided into four word-familiarity ranks. Twenty lists of 50 words are constructed in each rank by considering phonetic balance. Moreover, to compensate slight differences in intelligibility scores among word lists within the same word-familiarity rank, the relationship between the difference of word familiarity and signal-to-noise ratio (SNR) is estimated. Then this relationship is used to compensate the difference of the intelligibility score. The other compensation method is based on speech recognition threshold (SRT). The intelligibility score of each word can be equalized by changing the SNR according to the difference of the SRT of the word. The effectiveness of these compensation methods is discussed.

9:15

5aAA6. Binaural and spatial factors affecting speech intelligibility in rooms. H. Steven Colburn, Suzanne Carr, and Rui Wan (Hearing Res. Ctr. and Dept. of Biomed. Eng., Boston Univ., 44 Cummington St., Boston, MA 02215, colburn@bu.edu)

Prominent complaints of hearing-impaired listeners relate to the difficulties they experience in complex acoustic environments, particularly environments with multiple sources and reverberation. Available data and some theoretical ideas about the nature of the difficulties with speech intelligibility in these environments will be reviewed with impaired listeners in mind. The effects of the spatial characteristics of the sound sources and the room on the binaural signals and their processing is the primary focus of this presentation. Listeners with normal hearing as well as listeners with hearing impairments, with and without hearing aids, will be considered. Theoretical topics to be discussed include the effects of reverberation on interaural differences and on temporal modulation patterns, including the consequences for performance as predicted by standard models of binaural processing. [Work supported by NIH R01 DC00100.]

9:35

5aAA7. The effect of noise on novel word learning in sequential bilingual children. Pui Fong Kan, Kathryn Kohnert, and Peggy Nelson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

Preschool children must learn in noisy environments (Picard and Bradley, 2001), but young children are more negatively affected by background noise than are adults (Elliott, 1979). Young second-language learners are more negatively affected by noise than are monolingual children (Crandell et al., 1996; Nelson et al., 2005). Background noise may be still more detrimental in the case of preschool children who speak one language (L1) at home and who start to learn a second language (L2) in nursery school. In this present study we examine the effect of noise on fast-mapping skills in L1 (Hmong) and L2 (English). In the fast-mapping task each child was briefly presented with 5 novel words–counterbalanced in Hmong and in English–in a quiet setting over two different sessions. Another 5 novel Hmong and English words were presented to each child in two sessions in the presence of babble background noise. During each session the children were tested immediately after their exposure to the novel words. The results confirmed (1) that the children demonstrated–in both L1 and L2–better skills in quiet than in noise and (2) that the childrens fast-mapping skills were positively correlated with their pre-existing L1 and L2 language skills.

10:15

5aAA8. The effects of fluctuating interaural cues on the segregation of speech in rooms.   Douglas Brungart and Nandini Iyer (Air Force Res. Lab., 2610 Seventh St., Bldg 441, WPAFB, OH 45433, douglas.brungart@wpafb.af.mil)

Spatial separation is known to improve the segregation of talkers in anechoic environments, but relatively little is known about the role spatial cues play in speech segregation in reverberant rooms. One might expect the random disruptions in the interaural time and level differences (ILDs and ITDs) that occur in reverberant environments to eliminate many of the intelligibility benefits that normally occur for spatially separated speech. However, the precedence effect often leads to a robust perception of spatial location even in extremely echoic environments. This can result in an apparent separation between talkers that may facilitate selective attention to the location of the target speech even in a highly reverberant room. Also, random fluctuations in ITD and ILD may lead to differences in the apparent source widths of the target and masking sounds, which might serve as a segregation cue. In this talk, we examine the effects fluctuating interaural difference cues have on the segregation of target speech from competing speech or noise. The results suggest that differences in apparent source width can be used to segregate competing speech signals even when the target and masking signals appear to originate from the same direction relative to the listener.

10:35

5aAA9. The intelligibility advantage of clear speech in noise and reverberation backgrounds: Effects of speaking rate.   Jean C. Krause   (Commun. Sci. and Disord. Dept., Univ. of South Florida, 4202 E. Fowler Ave., PCD 1017, Tampa, FL 33617, jkrause@cas.usf.edu)

Clear speech refers to a speaking style that is significantly more intelligible than conversational speech for a variety of listeners and backgrounds, including roomlike backgrounds such as noise and reverberation [Payton *et al.*, J. Acoust. Soc. Am. **95**, 1581–1592 (1994)]. Although typically spoken more slowly than conversational speech [e.g., Uchanski *et al.*, J. Speech Hear. Res. **39**, 494–509 (1996)], talkers can produce clear speech at normal rates with training [Krause and Braida, J. Acoust. Soc. Am. **112**, 2165–2172 (2002)]. Whether this form of clear speech (clear/normal speech) provides similar intelligibility benefits to clear (clear/slow) speech in all situations has not been fully characterized, though some results [Krause and Braida, Iranian Audiol. **2**, 39–47 (2003)] suggest that the amount of benefit is more dependent on talker and environment. To investigate this possibility for roomlike degradations, the intelligibility advantage provided by clear/slow and clear/normal speech was evaluated by groups of eight normal-hearing listeners at five signal-to-noise ratios (-4, -2, 0, $+2$, and $+4$ dB) and three levels of reverberation (0, 0.18, and 0.6 s). Results for four talkers will be reviewed in order to illustrate the effects of speaking rate and speaking style on speech intelligibility in rooms.

10:55

5aAA10. Separating and understanding a talker from a mixture in reverberant spaces.   Barbara G. Shinn-Cunningham, Madhusudana Shashanka, and Scott Bressler   (677 Beacon St., Boston, MA 02215)

Many studies show that reverberant energy can interfere with the ability to understand a listener by smearing out the temporal modulations and distorting the spectral content that convey speech meaning. However, the amount of distortion caused by the reverberation in everyday settings often is not severe enough to degrade speech understanding when there is only one talker in a room. Despite this, even modest reverberation can have an enormous impact on the ability to understand the same talker when there is another sound source in the environment. One factor that likely contributes to the problem of understanding a talker in a mixture of sounds in reverberant spaces is that it is difficult to separate the target talker from the competing sounds. This talk will explore the ways in which reverberant energy degrades the ability to separate sounds in a mixture. The effects of reverberation on different cues for source segregation, including pitch and location, will be discussed. [Work supported by AFOSR and NIH.]

11:15

5aAA11. Preprocessing speech against reverberation.   Takayuki Arai   (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan)

Although short reverberation in a room may help speech perception, it is known that long reverberation causes degradation in speech comprehension. This is especially true for elderly people, the hearing-impaired, and non-native listeners. In order to prevent intelligibility degradation, we can apply a signal processing technique to the public address system of a room before speech signals are radiated through the loudspeakers. So far, we have developed several such preprocessing techniques. The two main techniques are modulation filtering [Kusumoto *et al.*, Speech Commun. **45**, 101–113 (2005)] and steady-state suppression [Arai *et al.*, Acoust. Sci. Technol. **23**, 229–232 (2002)]. Both techniques essentially enhance the temporal dynamics of speech between 1 and 16 Hz [Arai *et al.*, J. Acoust. Soc. Am. **105**, 2783–2791 (1999)]. Steady-state suppression, which suppresses steady-state portions of speech, reduces overlap-masking and improves speech intelligibility for young, elderly, and non-native listeners in reverberant environments [e.g., Hodoshima *et al.*, J. Acoust. Soc. Am. **119**, 4055–4064 (2006)]. This technique is more effective when speech-rate slowing is applied in advance [Arai *et al.*, Acoust. Sci. Tech. **26**, 459–461 (2005)]. Intelligibility with this technique exceeds a simple speech-rate slowing approach. [Work partially supported by JSPS.KAKENHI (16203041).]

5a SAT. AM

5aAA12. Failure of the precedence effect in children and in persons with hearing loss. Ruth Litovsky (Univ. of Wisconsin Waisman Ctr., 1500 Highland Ave., Madison, WI 53705)

Children spend a vast majority of time in reverberant environments. However, their auditory system appears not to be as fully equipped as that of adults for coping with reverberation. In our studies we simulate simple aspects of reverberation by exploiting a well-known auditory illusion known as the precedence effect (PE). Stimuli consist of clicks or speech, pairs of source (lead), and its first-arriving reflection (lag). Results from typically developing normal-hearing children ages 3–10 suggest that the PE undergoes significant maturation during childhood. In addition, we studied 10-year-old children with specific language impairment (SLI), who are delayed on expressive and receptive language. Performance was similar to that of younger, language-age-matched children, suggesting that the PE could be a useful clinical tool for measuring nonverbal temporal information processing. The PE was also used to study deaf adults with bilateral cochlear implants. Results suggest that binaural sensitivity is only available when the cochlea can be directly stimulated, but in free field the temporal smearing caused by the microphone leads to degraded and abnormal PE. Taken together, these studies suggest that a basic and necessary auditory function, that of suppressing echoes, fails under numerous circumstances, in children and in deaf, bilateral cochlear implant users.

SATURDAY MORNING, 2 DECEMBER 2006　　　　　　　　LANAI ROOM, 8:00 A.M. TO 12:00 NOON

Session 5aAB

Animal Bioacoustics: Animal Bioacoustics Poster Session (Poster Session)

Sheryl L. Coombs, Cochair
Bowling Green State Univ., Dept. of BioScience, Life Science Bldg., Bowling Green, OH 43403

Yutaka Hosokawa, Cochair
Ryukyu Univ., Dept. of Physiology, 207 Uehara, Nishihara-chu, Okinawa 903-0215, Japan

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon.

5aAB1. Reconstruction of bat sonar beams using a multichannel microphone array. Peter Stilz, Wiebke Pflästerer, and Hans-Ulrich Schnitzler (Tierphysiologie, Universität Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany, peter.stilz@uni-tuebingen.de)

We reconstruct the sonar beams emitted subsequently by flying bats from recordings of a two-dimensional microphone array. Based on the bats' position in space, which can be accessed through a infrared 3D-video system, and the simultaneous recordings of 16 small microphones, the sonar beams emitted by the bats and the sonar footprints on arbitrary planes can be estimated. For the interpolation of a continuous sonar beam at different frequency bands, directional frequency responses, atmospheric and geometric attenuation, and different lag times are taken into account for each microphone. Mean and maximum call intensities can be analyzed and various analysis parameters can be adjusted, different visualizations are available, and numerical data are exported. The system provides access to estimations of the maximum intensity of each call in front of the bat, beam width, shape, and intensity for arbitrary frequency bands, the aiming of the beam, and the bat's scanning behavior on its flight path.

5aAB2. A new low-power acoustic bat detector for long-duration observations. Matt Heavner (Univ. of Alaska Southeast, 11120 Glacier Hwy., Juneau, AK 99801)

A new small, low-power, long-duration, field-deployable computer for the acoustic monitoring of bats in Southeast Alaska has been developed. Parker *et al.*, (1997) report the northern geographic limit of four bat spe-

cies to be Southeast Alaska. (A fifth species, *E. fuscus*, goes much farther north.) The relationship of the bats to forest management practices, the habitat usage of the bats, and population size and trends are all very poorly known for bats in Southeast Alaska. Long duration monitoring of several different types of area (such as old-growth versus recently logged forest) will provide knowledge to improve management practices in regards to bat ecology in Southeast Alaska. With the motivations just described, the hardware, development methods, and analysis software designed to develop an improved detector are presented. Tradeoffs between computation power, power consumption, cost, and other factors will be discussed. The project has benefitted in several direct ways from the system design requirement of modularity. [Work supported by Alaska epartment of Fish & Game.]

5aAB3. Development of networked weatherproof remote sensing embedded Linux system for bioacoustical evaluation. Hiroki Kobayashi, Yoshihiro Kawasaki, and Shinichi Watanabe (Matsushima Lab., Dept. of Geosystem Eng., The Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan, tt56958@mail.ecc.u-tokyo.ac.jp)

The purpose of this research is to develop a real-time evaluation technology of bioacoustic diversity in a remote location by networked Linux embedded system. The hardware is designed to operate 24 h a day, 7 days a week, 365 days a year under severe environment, such as extremely high moisture and temperature environment in subtropical forest. Three stages of the hardware development were completed at this point. At first, a wired long-lived bioacoustic monitoring system in subtropical forest in Iriomote Island, Japan was developed and succeeded to operate continuously for the

past 2 years based on an unmanned condition with less than 1000 US dollars annual budget. Based on this development, a wired multifunctional monitoring system with infrared cameras and microclimate monitoring sensors was developed and evaluated with less than 4000 US dollars. At last, with FOMA (Freedom of Mobile Multimedia Access) mobile information technology offered by Japanese mobile phone operator NTT DoCoMo, a multifunctional bioacoustic monitoring system with two-way communication system which operates passive and active observation for Iriomote Cat was developed with less than 8000 US dollars. [Work supported by NTT DoCoMo.]

5aAB4. Can birds discriminate between simple sounds and natural vocalization with different degrees of reverberation. Sandra H. Blumenrath and Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742)

Birds in natural habitats using long-range acoustic signals experience substantial acoustic alterations of their signals when transmitted from a sender to a receiver. The main degradation comes from reflections of the emitted sounds from surfaces in and around the transmission pathway. These reverberations cause tails of echoes and affect frequency and amplitude modulation within the sounds' original time frame. Humans discriminate among reverberant and nonreverberant speech segments and use the nature of reverberation in determining the location of the sound source. Using operant conditioning and psychophysical methods, we investigated the ability of several species of small birds to discriminate between nondegraded sounds and sounds artificially reverberated to varying degrees, while holding other sound aspects constant. Birds were able to easily discriminate between nonreverberant sounds and sounds reverberated to various extents. By comparing the birds' performance on stimuli varying along different dimensions of reverberation, we can make predictions as to the birds ability to use reverberations as acoustic cues in distance estimation and locating sound sources. [Work supported by NICD DC000198 (to R.J.D.) and DC004664.]

5aAB5. Perception of budgerigar (*Melopsittacus undulatus*) warble elements. Hsiao-Wei Tu and Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742)

The warble song of budgerigar (*Melopsittacus undulatus*) contains various classes of vocal elements, including contact calls, alarm calls, harmonic calls, etc. Previous work has shown that budgerigars can discriminate different call types according to both acoustic and functional features; however, there is no evidence on how well birds can distinguish subtle differences within and between each category. Using operant conditioning and psychophysical methods, budgerigars were trained to discriminate among different warble elements. In general, birds were able to discriminate quite easily between all warble elements, even those with high spectral cross correlations. Humans, on the other hand, had difficulty distinguishing among warble elements that were similar to one another. The relative discriminability of warble elements within and across categories is used to more precisely define the functional categories of warble elements in understanding vocal communication in this species. [Work supported by NIDCD DC000198 to R.J.D. and P30 DC004664.]

5aAB6. Voices of the dead: Complex nonlinear vocal signals from the larynx of an ultrasonic frog. Roderick A. Suthers (School of Medicine and Dept. of Biol., Indiana Univ., Bloomington, IN 47405, suthers@indiana.edu), Peter M. Narins (Univ. of California, Los Angeles, CA 90095), Wen-Yu Lin (Univ. of Illinois, Urbana, IL 61801), Hans-Ulrich Schnitzler, Annette Denzinger (Univ. Tuebingen, D-72076 Tuebingen, Germany), Chun-He Xu (Chinese Acad. of Sci., Shanghai, P. R. China), and Albert S. Feng (Univ. of Illinois, Urbana, IL 61801)

Most anurans are highly vocal but their vocalizations are stereotyped and simple with limited repertoire sizes compared to other vocal vertebrates, due presumably to the limited mechanisms for fine vocal motor control. We recently reported that the call of the concave-eared torrent frog

(*Amolops tormotus*) is an exception in its seemingly endless variety, musical warbling quality, extension of call frequency into the ultrasonic range, and the prominence of nonlinear features such as period doubling. We now show that the major spectral features of its calls, responsible for this frog's vocal diversity, can be generated by forcing pressurized air through the larynx of euthanized males. Laryngeal specializations for ultrasound appear to include very thin portions of the medial vocal ligaments and the reverse sexual size dimorphism of the larynx being smaller in males than in females. The intricate morphology of the vocal cords, which changes along their length, suggests that nonlinear phenomena likely arise from complex nonlinear oscillatory regimes of separate elastically coupled masses. *Amolops* is thus the first amphibian for which the intrinsic nonlinear dynamics of its larynx, a relatively simple and expedient mechanism, can account for the species call complexity, without invoking sophisticated neuromuscular control.

5aAB7. Comparative analysis of head-banging behavior in the subterranean termites *Coptotermes formosanus* and *Reticulitermes flavipes* using high-speed imaging. Tom Fink, Lichuan Gui, Yong Wang, Zhonghua Cao (Natl. Ctr. for Physical Acoust., One Coliseum Dr., University, MS 38677), Adarsh Jaiswal, Orwa Tahaineh, Roger Hasse, John Seiner (Natl. Ctr. for Physical Acoust., University, MS 38677), and Alan Lax (U.S. Dept. Agriculture, Blvd. New Orleans, LA 70124)

Disturbed soldiers of *Coptotermes formosanus* (FST) readily head-bang (HB) only in their galleries and carton nests while *Reticulitermes flavipes* (Rf) will also readily HB in an artificial chamber. The head-banging motion was examined using the Photron FASTCAM-ultima APX video camera at 10 000 frames per second. Termites were placed in a 0.25-in. diameter hole in a wood block and induced to HB by touching them with a small brush. A PCB Piezotronics 352B accelerometer was used to detect HB-induced vibrations. These vibration signals were then received by the Photron Multi-Channel-Data-Link apparatus, which then superimposes the vibration waveform on the video frames corresponding to the head movement. Thus, movement of the head can be correlated with the vibrations produced. Time between successive HB is about 70 ms or more in FST vs 40 ms or less in Rf despite the much larger head of Rf. Upward and downward head velocities in FST are, respectively, 100 and 240 mm/s, while the respective values for *Reticulitermes* are over 200 and about 400 mm/s. Acceleration for FST is about 400 m/s or up to 700 m/s^2 for *Reticulitermes*. Both species may show head rebounding after the initial contact with the substratum.

5aAB8. Underwater auditory localization by harbor seals (*Phoca vitulina*). Anaïs Bodson, Lars Miersch, and Guido Dehnhardt (General Zoology & Neurobiology, Univ. of Bochum, ND 6/33, D-44780 Bochum, Germany, bodson@marine-science-center.de)

The underwater sound localization acuity of a swimming harbor seal was measured in the horizontal plane. The stimulus was either a double sound (two 6-kHz pure tones lasting 0.5 s separated by an interval of 0.2 s) or a single continuous sound of 1.2 s. Testing was conducted in a 10-m-diameter underwater half-circle with hidden loudspeakers. The animal was trained to swim along the diameter of the half-circle and to change its course towards the sound source right after the signal was given. The seal indicated the sound source by touching its assumed position at the half-circle. The deviation of the seal's choice from the actual sound source was measured by means of video analysis. In trials with the double sound the seal localized the sound sources with a mean deviation of 2.8° and in trials with the single sound with a mean deviation of 4.5°. In a second experiment minimum audible angles of the animal were determined as a function of frequency (2, 4, 6, 8, and 16 kHz resulting in MAAs of 9.4°, 9.6°, 9.8°, 13.0°, and 13.5°, respectively). Intraspecific differences of the MAA will be presented for five additional harbor seals.

5a SAT. AM

5aAB9. Evaluation of the auditory capabilities of marine mammals using a portable auditory-evoked potential system. Kristen A. Taylor, Paul E. Nachtigall, T. Aran Mooney, Michelle M. Yuen (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua HI 96734), and Alexander Ya Supin ((Russian Acad. of Sci., Moscow, Russia)

There is growing concern about the increasing amounts of acoustic pollution in our oceans. Anthropogenic noise sources span a large range of frequencies and amplitudes. The extents to which these noise sources adversely impact marine mammals is poorly understood yet they are of particular significance given that the acoustic environment is crucial for many aspects of marine mammal life. Past and present experiments have addressed the effects of different sounds on the hearing capabilities of captive dolphins. However, only a few individuals from a small number of species have been tested. Our research group has created a portable system that is capable of measuring the hearing thresholds of marine mammals using auditory-evoked potential (AEP) techniques. This portable system has enabled us to obtain audiograms of otherwise inaccessible marine mammal species. This system will also allow us to measure the hearing of stranded animals as well as begin to quantify the normal variations in hearing between different ages and genders of animals in captivity.

5aAB10. Optical imaging of neural activity in the right and left guinea pig auditory cortices evoked by repeated sound. Yutaka Hosokawa (Dept. of Physiol., Ryukyu Univ., Nishihara-tyo, Okinawa, 903-0215 Japan) and Junsei Horikawa (Toyohashi Univ. of Technol., Toyohashi, 441-8580 Japan)

Processing of spectral and temporal information in the right and left auditory cortices of guinea pigs anesthetized with ketamine and xylazine was investigated using an optical imaging technique with a voltage-sensitive dye (RH795). Tonotopical organization in the primary (AI) and the dorsocaudal (DC) fields in the left and the right auditory cortices were compared, as visualized by superimposition of the areas activated by tones at 2, 4, 8, and 16 kHz. In 65% of 31 animals, the distance between 2- and 16-kHz isofrequency bands was longer in the left than in the right AI; it was shorter in the left than in the right DC. Repetition rate transfer functions (RRTFs) in AI measured with repeated sounds (420 Hz repetition rates) were low pass, with a sharp drop-off above 10 Hz. In contrast, RRTFs in DC were bandpass with a peak at 8 or 10 Hz. The cutoff frequencies of RRTF were different for each of the left and right cortices.

5aAB11. The physoclistous swim bladder of chaetodontid butterflyfishes: Implications for acoustic function. Christopher Woods, Jacqueline Webb (Dept. of Biol., Villanova Univ., Villanova, PA 19085; jacqueline_webb@mail.uri.edu), and Darlene Ketten (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Butterflyfishes (genus *Chaetodon*) have a swim bladder lateral line connection (laterophysic connection, LC), hypothesized to convert sound-induced oscillations of the swim bladder into fluid flow in the lateral line system and/or ear. Evaluation of LC function is dependent upon an understanding of swim bladder acoustics, which is a function of swim bladder anatomy. We used several anatomical methods (including CT) to describe swim bladder and tunica externa and tunica interna morphology (including gas gland/*rete mirabile* complex and the oval, responsible for gas secretion and resorption) in *Chaetodon* and *Forcipiger*. Swim bladder and tunica externa morphology differ between *Chaetodon* spp. with different LCs morphologies. A perforated transverse diaphragm divides the gas volume into two compartments. The gas gland/*rete mirabile* complex is in the ventral midline of the anterior compartment; it varies in morphology and is largest in species with a direct LC. The oval, defined by an extensive capillary network, occupies the entire tunica interna of the posterior compartment, so that diaphragm position determines oval size. These data raise questions about swim bladder structure-function relationships with respect to the reception and transduction of acoustic stimuli in coral reef fishes. [Work supported by NSF Grant IBN-0132607 to J.F.W.]

5aAB12. The temporal resolution of goldfish hearing: An auditory evoked potential study of gap detection. Jianqiang Xiao and Christopher Braun (Dept. of Psych., Hunter College, 695 Park Ave., New York, NY 10021, jxiao@hunter.cuny.edu)

Temporal processing in goldfish (*Carassius auratus*) was studied by measuring auditory evoked potentials (AEP) using gaps in continuous band-limited Gaussian noise. Long (>100 ms) silent gaps in high amplitude noise (30 dB SL) evoked distinct offset and onset responses, lasting 6090 ms. Offset and onset responses overlapped with shorter gaps and became difficult to distinguish for gaps <10 ms. The gap-response waveform was modeled as the sum of offset and temporally-shifted onset responses from the same animals (using 120-ms gaps). This model predicted the waveform of gap responses nearly perfectly for gaps longer than 6 ms. Waveforms evoked by shorter gaps (<6 ms) did not accurately fit this model, suggesting that one or both components of the gap response were inhibited or altered. Nonetheless, clear responses were evoked by gaps shorter than 1 ms at this intensity, while longer gaps (up to 10 ms) were required for detection in a low-intensity noise background. This study extends and confirms prior reports that the temporal resolution for goldfish is on the order of 1 ms, but there is also evidence for longer integration processes, over a period of several ms. [Work supported by Grants S06GM60654, R03MH067808, RR03037 (NIH).]

5aAB13. The effects of source distance and orientation on lateral line stimulation patterns and source localization by Lake Michigan mottled sculpin. Sheryl Coombs and Paul Patton (Dept. of Biological Sci., JP Scott Ctr. for Neurosci., Mind, and Behavior, Bowling Green State Univ., Bowling Green, OH 43403)

Visually deprived, Lake Michigan mottled sculpin respond to small dipole sources (sinusoidally vibrating spheres) with an initial orienting response and subsequent approach and strike behaviors. Lateral line canal neuromasts, but not superficial neuromasts, are required for the initial orienting response. Computational (potential flow) models reveal that stimulation patterns across lateral line canals vary in complex ways when the orientation (axis of vibration) and distance of the dipole source are varied. The body location where lateral line stimulation is maximal (the excitatory peak) might seem to provide a reliable cue of source location. However, this cue is (1) confounded by the relative amplitude, number, and body location of excitatory peaks when source orientation is varied and (2) degraded by both an increase in peak width and a reduction in peak amplitude when source distance is increased. Behavioral measures of orienting accuracy for equally detectable sources at two distances (3 and 6 cm) and horizontal orientations (parallel or orthogonal to the long axis of the fish) reveal that localization abilities are degraded by increases in source distance. Although varying source orientation does not produce dramatic deficits in orienting accuracy, more subtle effects were observed that may relate to information encoded in spatial excitation patterns.

5aAB14. The use of multimodal communication during courtship in Malawi cichlids. Adam R. Smith and Moira J. van Staaden (Dept. of Biological Sci. and J. P. Scott Ctr. for Neurosci., Mind & Behavior, Bowling Green State Univ., Bowling Green, OH 43403)

The cichlids of Lake Malawi are often considered the quintessential example of a species radiation driven primarily by sexual selection and

relying on visual cues. However, the description of sound production by courting males in several species raises the possibility that acoustic signaling may play a significant role in the selection process. This study describes the mating calls from multiple species in the genera *Melanochromis* and *Metriaclima* in order to discern if there is active selection of call characteristics between closely related species or whether differences are more likely the result of drift. Variation in total call duration and fundamental frequency within genera often exceeded that between genera, particularly between the sister species *Melanochromis johanni* and *Melanochromis cyaneorhabdos*. This suggests that mating calls may indeed be subject to divergent sexual selection in closely related species. Moreover, *in situ* sound recordings from Lake Malawi indicate that this behavior is of ecological relevance and not simply an artifact of laboratory conditions. Together, these results suggest that current models of sexual selection in Malawian cichlids may benefit from consideration of previously ignored sensory modalities. The plasticity inherent in such multimodal processing may prove of particular importance to the species-rich African cichlid radiation.

5aAB15. Reaction of plaice (*Pleuronectes platessa*) and cod (*Gadus morhua*) to low-frequency sound in an experimental tank. Christina Mueller-Blenkle (FRS Marine Lab., P.O. Box 101, Aberdeen AB11 9DB, UK) and Karin Luedemann (Institut fr Angewandte Fischbiologie, 22083 Hamburg, Germany)

Anthropogenic sound in the sea is an increasing problem and more knowledge on the influence of sound on fish needs to be collected. For this reason experiments are carried out in an experimental tank 10 m in diameter to examine the reaction of juvenile and adult cod and plaice to sound. Pure tones from the frequency range of offshore wind turbines are artificially emitted in typical wind turbine sound levels of 130 and 140 dB *re* 1 Pa. A sound gradient in the tank enables the fish to avoid highest sound levels. The number of fish in the vicinity of the sound source is determined in regular time intervals before, during, and after the 24-h period of sound exposure. Changes in numbers of fish can be seen in the periods with and without sound in many experiments. Reactions of plaice are ambiguous, pointing to both avoidance of or attraction to the sound. In the majority of the experiments carried out on cod, avoidance behavior to the sound can be seen with strongest reactions at 25, 60, and 90 Hz. The results indicate potential avoidance behavior of cod to the operational noise of offshore wind farms.

SATURDAY MORNING, 2 DECEMBER 2006   KAHUKU ROOM, 9:00 TO 11:10 A.M.

Session 5aBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Contrast Agents

Charles C. Church, Cochair

Univ. of Mississippi, National Center for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—9:00

Contributed Papers

9:05

5aBB1. Microbubble spectroscopy of ultrasound contrast agents. Sander van der Meer, Benjamin Dollet (Phys. of Fluids, Univ. of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands), Chien T. Chin, Ayache Bouakaz, Marco Voormolen, Nico de Jong (Erasmus MC, 3000 DR Rotterdam, The Netherlands), Michel Versluis, and Detlef Lohse (Univ. of Twente, 7500 AE Enschede, The Netherlands)

We present a new optical characterization of the behavior of single ultrasound contrast bubbles. The method consists of insonifying individual bubbles several times successively sweeping the applied frequency and recording movies of the bubble response up to 25 million frames per second with an ultra-high-speed camera operated in a segmented mode. The method, termed microbubble spectroscopy, enables one to reconstruct a resonance curve of the oscillating bubble in a single run. We analyze the data through a linearized model for coated bubbles. The results confirm the significant influence of the shell on the bubble dynamics: shell elasticity increases the resonance frequency by about 50%, and shell viscosity is responsible for about 70% of the total damping. The obtained value for shell elasticity is in quantitative agreement with previously reported values. The shell viscosity decreases significantly with the dilatation rate, revealing the nonlinear behavior of the phospholipid coating.

9:20

5aBB2. High-frequency response of polymer shell contrast agents. Jeffrey A. Ketterling, Jonathan Mamou (Lizzi Ctr. for Biomed. Eng., Riverside Res. Inst., 156 William St., New York, NY 10038), John S. AllenIII (Univ. of Hawaii-Manoa, Honolulu, HI 96822), Orlando Aristizábal, Rene G. Williamson, and Daniel H. Turnbull (NYU School of Med., New York, NY 10016)

Ultrasound contrast agents were originally designed for use at low MHz frequencies; however, recent experimental evidence suggests that many of these agents may also be useful for contrast imaging at frequencies above 20 MHz. Three polymer shelled agents (mean diameters of 0.56, 1.1, and 3.4 μm) from POINT Biomedical were investigated to determine their optimal high-frequency responses. A flow phantom was constructed to restrict the flow of a dilute contrast agent solution to a small volume. The focus of a 40-MHz single-element transducer from a Visual Sonics V770 was positioned within the flow phantom. Radio-frequency backscatter data from individual contrast agents were digitized under a variety of pulse durations (1–15 cycles) and pressure levels (1.4–7.6 MPa). Echo signals from individual bubbles were windowed, and spectra were calculated. Each of the POINT agents reflected energy at the 40-MHz fundamental; however, only the 1.1-μm-diameter agent displayed characteristics of a harmonic response under longer excitations (15 cycle) and medium pressure settings (2.5 to 5 MPa). Experimental results were com-

pared to theoretical calculations for the case of the polymer shell remaining intact during excitation and for free gas nuclei that result from a ruptured shell.

9:35

5aBB3. Optical observations of the collapsing behavior of polymer micro-capsule irradiated in the ultrasonic sound field. Tetsuo Yamato, Kenji Yoshida, Wataru Kiyan, and Yoshiaki Watanabe (Faculty of Eng., Doshisha Univ., 610-0321 1-3 Tataramiyakotani Kyotanabe Kyoto Japan)

For the medical use, it is important to understand the micro-capsule behavior during collapsing under the ultrasonic sound field. In this report, the behavior of micro-capsule collapse was observed using the high-speed video camera system (HPV-1, SHIMADZU: maximum recording rate; 1 000 000 frames/second). By using this system, the detail phenomena can be observed. The micro-capsule leaked the internal gases through a small pin hole on the surface of the micro-capsule while repeating asymmetrical vibration. The internal gases were emitted during the positive period of driving sound pressure and returned again in the micro-capsule during the negative period. By repeating the outflow and inflow of internal gases, the stress was cumulated on the shell. Therefore, the small pin hole developed into the big one, then, the micro-capsule emitted a large amount of internal gases and the capsule was collapsed. After the collapsing, only the shell was remained. From the results, it was found for the effect of the shell, that is, the micro-capsule radius shrunk in positive pressure phase, however it hardly expanded in negative pressure phase. These phenomena were numerically simulated by the vibration model, which the effect of the shell was removed only when shrinking. This simulation results qualitatively agreed with observation results.

9:50

5aBB4. Thermal effect of microbubbles in the focused ultrasound field. Yukio Kaneko, Naoyuki Iida, Shu Takagi, and Yoichiro Matsumoto (Dept. of Mech. Eng., The Univ. of Tokyo, Tokyo, Japan)

High-intensity focused ultrasound (HIFU) has been developed for the treatment of tumor, and additionally the medical applications with microbubbles such as ultrasound imaging have attracted much attention. In the field of the bubble dynamics, it is known that the microbubble is an energy converter of ultrasound mechanical energy to heat when subjected to an acoustic field. The goal is the tissue heating enhanced by microbubbles. In this study, the relationship between the heating effect and the behavior of microbubbles is analyzed. The temperature rise was measured by a thermocouple and a thermal liquid crystal sheet, and the bubble behavior was simultaneously observed with a camera. The ultrasound frequency is 2.2 MHz and the intensity is 100–2000 W/cm^2. The temperature rise became larger as the number density of microbubbles (Levovist) around the focal region increased. As the number density became too large, the conversion efficiency from ultrasound energy to heat became smaller because the shielding effect of ultrasound by the existence of too many bubbles became larger. From these results, it is indicated that it is important to choose the optimal conditions of microbubbles and ultrasound in order to control the heat deposition from microbubbles.

10:05–10:25   Break

10:25

5aBB5. Engineering the acoustic response of lipid-based microbubbles: Simulations and high speed optical imaging results. Eleanor Stride (Dept. of Mech. Eng., UCL, Torrington Pl., London WC1E 7JE, UK, e_stride@meng.ucl.ac.uk), Rob Eckersley, Kevin Chetty, Charles Sennoga, and Jo Hajnal (Imperial College, London, W12 0NN, UK)

Gas microbubbles coated with a surfactant or polymer shell have become well established as contrast agents for ultrasound imaging and are under investigation for therapeutic applications such as targeted drug delivery. The purpose of this study was to examine the effect of modifying the microbubbles' coating composition upon their response to ultrasound excitation. Using a high speed camera operating at 3 MHz, the radial oscillations from the commercial contrast agent SonoVueTM and four in-house microbubble preparations were measured under controlled insonation conditions (four-cycle Gaussian pulse; center frequency 0.5 MHz; peak negative pressure 30–70 kPa). The results indicate that microbubble response varies as the lipid structure and functional components in the coating change. These findings have been compared with simulation results using the generalized model for microbubble dynamics developed by the authors, in order to quantify the observed effects of changing the shell composition upon the microbubbles viscoelastic and diffusion characteristics. Both microbubble shrinkage and variable shell viscoelasticity can have significant implications for the acoustic response. This in turn can affect contrast specific imaging methods, especially those involving multipulse sequences such as pulse inversion. Similarly, these phenomena may affect the microbubble destruction threshold and hence the efficiency of drug delivery procedures.

10:40

5aBB6. The use of homemade microbubbles in detection of organ bleeding and enhancement of high-intensity focused ultrasound therapy. Wenbo Luo, Vesna Zderic, and Shahram Vaezy (Dept. of Bioengineering, Univ. of Washington, Seattle, WA 98195)

We investigated the feasibility of using homemade microbubbles in detection of internal bleeding, and for optimization of high-intensity focused ultrasound (HIFU) therapy. Homemade microbubbles were produced by sonicating albumin and dextrose saline solution with the presence of perfluorocarbon gas. Microbubbles were injected intravenously to reveal rabbit kidney and liver injuries. HIFU treatments were performed on liver lacerations for hemostatis and on intact liver for lesion production, with or without homemade microbubbles. The mean size of homemade microbubble was 4.7 um (range 1.2–7.1 μm). The concentration of microbubbles was 3.1*108 bubble/ml. Kidney and liver injuries were revealed by mosaic color flow at the injury sites. In treatment, the average hemostasis times normalized with the bleeding rate were 39 s/ml*min for HIFU only and 21 s/ml*min for HIFU+microbubbles treatments. The presence of microbubbles reduced 44 percent ($p<0.05$) time in producing the first formation of the blood coagulum. The HIFU lesion depth was reduced by 35 percent because microbubbles shield the postfocal region from HIFU exposure. Homemade microbubbles oscillate and collapse in ultrasound field, generating mosaic pattern thus revealing the bleeding site. It also shows potential in promoting HIFU hemostasis and shielding deep-tissue regions from HIFU exposures.

10:55

5aBB7. The effects of encapsulation on growth and dissolution of a contrast microbubble. Kausik Sarkar and Pankaj Jain (Mech. Eng., Univ. of Delaware, Newark, Delaware, sarkar@me.udel.edu)

Micron-size gas bubbles are intravenously injected into patients body at the time of ultrasound imaging to improve image contrast. The bubbles are encapsulated by a thin layer (4–10 nm) of protein, lipids, and other surface active materials, to prevent their premature dissolution in the blood. We will present a model for the dissolution of the microbubble that accounts for the effects of encapsulation. The encapsulation hinders the permeability of the gas-liquid surface and its elasticity balances the surface tension-induced stress. Both these effects will be explicitly modeled. The model behavior will be discussed for variations of the material parameters and conditions (encapsulation permeability and elasticity, mole fraction of the osmotic agent and liquid saturation). The encapsulation significantly affects the bubble growth and dissolution including their time scales.

Session 5aEA

Engineering Acoustics: Development and Application of Practical Electroacoustical Devices

Juan I. Arvelo, Cochair

Johns Hopkins Univ., Applied Physics Lab., 1100 Johns Hopkins Rd., Laurel, MD 20723-6099

Ryo Mukai, Cochair

NTT Communication Science Labs., 2-4 Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

Contributed Papers

8:00

5aEA1. Wideband piezoelectric rectangular loudspeaker using a tuck-shaped polyvinylide-fluoride bimorph. Taira Itoh, Juro Ohga (Shibaura Inst. of Technol., 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan, m106011@shibaura-it.ac.jp), Toshitaka Takei (Take T Co.), and Nobuhiro Moriyama (Kureha Co.)

A bimorph sheet of polyvinylide-fluoride (PVDF) film was applied to a flat rectangular loudspeaker as a folded zigzag-tack shape diaphragm whose size is, e.g., 260–144 mm with various depths. These loudspeakers, which we call bimorph tuck loudspeakers, are characterized by their moderate size, wide frequency range, light weight, and lack of magnetic flux radiation. This study examines the electro-acoustic transducer characteristics of these loudspeakers. Their sensitivity and resonance frequency were measured using a flat-panel baffle. Their electrical properties and sound field characteristics were also examined. A speaker's resonance frequency was estimated theoretically as a function of tuck depth. The measured resonance frequencies were, for example, 200, 160, 90, and 60 Hz. These low resonance frequencies are satisfactory for full-range loudspeaker use. Frequency characteristics of output sound pressure were also measured. They still include some irregularities that are attributable to a local bending resonance of the diaphragm. This paper examines the control methods for those irregularities.

8:15

5aEA2. Digital measurement method for dynamic distortion of loudspeakers. Keiichi Imaoka and Juro Ohga (Shibaura Inst. of Technol., 3-7-5 Toyosu, Koutou-ku, Tokyo, 135-8548, Japan, m605101@sic.shibaura-it.ac.jp)

Though many sorts of measuring methods using digital signal processing techniques have been developed, most of them are only for measurements in a linear range. There is still no suitable digital measuring method for nonlinear distortion of acoustical devices. An accurate and convenient nonlinear distortion measurement system should be developed for evaluation of electroacoustical transducers. A new digital distortion measuring method for acoustical devices is presented briefly. This method applies a Pink-TSP signal (time stretched pulse, i.e., quickly swept sinusoidal signal), whose frequency band is partially eliminated, to an acoustical system to be measured. The detected component produced in the rejected band is picked up by a bandpass filter and measured as a distortion. First, two kinds of small size electrodynamic loudspeaker units were measured by using the dynamic distortion measurement system mentioned above. Next, distortion characteristics of mechanical single resonant system models whose component shows amplitude-dependent nature were examined. This paper describes the relationship between distortion characteristics and quality factor, input signal level, and nonlinear parameters. The difference in distortion of two kinds of loudspeakers is detected by this measurement. It is seen that the resonant phenomenon has an influence on distortion characteristics.

8:30

5aEA3. Effect of acoustical load parameters for frequency response of headphones. Megumi Kobayashi, Juro Ohga, Wataru Onoda (Shibaura Inst. of Technol., 3-7-5 Toyosu, Koto-ku, Tokyo, 135-8548 Japan), and Ikuo Oohira (Ashida Sound Co., Ltd. Shinagawa-ku, Tokyo, 141-0032 Japan)

This report describes experimental and theoretical analyses for characteristics of headphones to examine the effect of significant parameters for design. Though various types of headphones are being used, the standard design procedure is still not established, because physical relationship between headphone construction and the human ear is not understood satisfactorily. The authors measured frequency response of various headphones by using IEC 60318 artificial ear and IEC 60959 head and torso simulator (HATS), and also measured the response by human ears by inserting a miniature microphone for comparison. By these examinations, a few mechanoacoustical parameters important for design were picked out. Especially, the effect of acoustical leakage shall be noted. Design parameters for supra-aural and intraconcha headphones will be reported.

8:45

5aEA4. High-output microsized earplug driver system for extreme acoustic environments. Keehoon Kim, Andrew Kostrzewski (Physical Optics Corp., 20600 Gramercy Pl., #100, Torrance, CA 90501, kkim@poc.com), and Reginald Daniels (Air Force Acoust. Res. Lab., Wright–Patterson AFB, OH 45433)

Hearing protection is critical when aircraft flight operations create a high-noise environment (130–150 dB) for aircrew and service members. Untreated aircraft noise causes noise-induced hearing loss (NIHL) and poses risks of similar hearing damage in the future. The authors have developed a new microsized smart material actuator (MSMA), a high-output earplug driver for active noise reduction (ANR) hearing protection, based on highly efficient microstructured acoustic exciters. The MSMA consists of a microscale earplug (MEP) with a proprietary acoustic structure, a protective MEP package, and a compact actuation amplifier easily connected to an external ANR controller. Performance tests show that the system produces over 130 dB into a 1-cc trapped cylindrical volume, 6 mm in diameter by 7 mm long, without significant phase delay or noticeable discontinuities in frequency response.

9:00

5aEA5. The surface and bulk microfabrication of optical seismometers and vibrometers using Sandia National Laboratories' silicon micromachining technology. Murat Okandan, Neal A. Hall, Robert Littrell (Sandia Natl. Labs., Albuquerque, NM 87185), Baris Bicen, and F. L. Degertekin (Georgia Inst. of Technol., Atlanta, GA 30332)

The microfabrication of optical seismometer and vibrometer structures using Sandia's silicon-based process technology is presented. The structure employs a 1.5-mm-diameter polysilicon diaphragm anchored to a

5a SAT. AM

3329   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3329**

200- to 400-μm-thick bulk silicon proof mass, and suspended by surface micromachined polysilicon springs. Motion of the diaphragm is detected using an integrated polysilicon surface micromachined grating-based optical interferometer. Initial dynamic characterization of these structures in vacuum shows a fundamental resonant frequency of 250 Hz. Based on the high force sensitivity of these structures combined with interferometric motion detection, 1-μg/$\sqrt{\text{Hz}}$ detection resolution in the 1-Hz frequency range is estimated. Microfabricated structures employing electrostatic actuation electrodes for the diaphragm are also presented, which enable force feedback operation for high dynamic range and enhanced dc stability. These characteristics may make these sensors well-suited for seismic and inertial navigation applications.

9:15

5aEA6. Finite-element modeling of thin-film damping in micromachined microphones employing diffraction based optical readout. Neal A. Hall, Murat Okandan (Sandia Natl. Labs., Albuquerque, NM 87185), and F.L. Degertekin (Georgia Inst. of Technol., Atlanta, GA 30332)

Micromachined microphones with diffraction-based optical displacement detection have been presented in detail previously [J. Acoust. Soc. Am. **118**(5), 3000–3009 (2005)]. In addition to providing superior diaphragm displacement detection sensitivity over miniature capacitive sensors, optical approaches have the advantage of removing mechanical design constraints on the microphone backplate perforation architecture as well as constraints on the diaphragm-backplate gap thickness. Taking advantage of this freedom to tailor design the frequency response function and thermal noise characteristics of miniature high-performance microphones requires a sophisticated damping model to navigate this new design space. A finite-element model in ANSYS based on the modal projection method is employed to study the dynamics of new optical microphone structures. The model extracts the frequency-dependent resistance and stiffening characteristics of the film using modal displacement profiles of the diaphragm. Simulated frequency response functions and thermal-mechanical noise limits agree well with those measured on fabricated structures. Most notably, 1.5-mm-diameter diaphragm structures with under 1 μPa/$\sqrt{\text{Hz}}$ thermal noise and over 20-kHz bandwidth have been successfully designed, fabricated, and characterized. [The authors would like acknowledge the IC Postdoctoral Research Fellowship Program.]

9:30

5aEA7. Design optimization of a microelectromechanical piezoresisitive microphone for use in aeroacoustic measurements. Brian Homeijer, Ben Griffin, Toshi Nishida, Lou Cattafesta, and Mark Sheplak (Univ. of Florida, 231 MAE-A Bldg., Gainesville, FL 32611-6250, sheplak@ufl.edu)

A microelectromechanical systems (MEMS)-based piezoresistive microphone optimimum design is presented, with a focus on improving the minimum detectable pressure over many current technologies without sacrificing bandwidth. This microphone design addresses many of the problems associated with previous piezoresistive microphones. Here, a novel nonlinear circular composite plate mechanic model was employed to determine the stresses in the diaphragm, which was designed to be in the compressive quasibuckled state. With this model, the inherent in-plane stresses that occur in the microelectronic fabrication process can be used to increase the sensitivity of the device. Ion-implanted doped silicon was chosen for the piezoresistors and a fabrication recipe was made which minimizes the inherent noise characteristics of the material. The piezoresistors are arranged in a Wheatstone bridge configuration with two resistors oriented for tangential current flow and two for radial current flow. A lumped element model was created to describe the dynamic characteristics of the microphone diaphragm and the cavity/vent structure. The geometry for this device was optimized using a sequential quadratic programming scheme performed using the aforementioned novel device characteristics. Results indicate a dynamic range in excess of 120 dB for devices possessing resonant frequencies beyond 120 kHz.

9:45

5aEA8. A two-dimensional mechanical resonator array filter with reduced sensitivity to disorder. John A. Judge, Joseph F. Vignola (Mech. Eng. Dept., The Catholic Univ. of America, Washington, DC 20064, judge@cua.edu), and Scott A. Mathews (The Catholic Univ. of America, Washington, DC 20064)

The structural acoustics of perfectly periodic arrays of coupled mechanical resonators are characterized by frequency passbands, outside of which waves do not propagate across the array, but in which wave propagation is attenuated only by dissipation. Arrays of high-Q micromechanical resonators, with sufficiently high resonant frequencies, can thus be used as filters for rf band electrical signals. However, small deviations from periodicity, such as the disorder induced by finite fabrication tolerances, are known to have significant consequences for energy propagation across such arrays. This is particularly true in narrow passband applications, in which resonators are weakly coupled, and the passband energy transmission can be significantly degraded. In this work, a two-dimensional array of resonators is demonstrated that is less sensitive to disorder than equivalent one-dimensional arrays, improving filter performance without the need for improved manufacturing tolerances. The concept is explained in terms of a simple model, and the design of a prototype array is described, including a statistical investigation of disorder effects using finite-element analysis. Experimental measurements are reported demonstrating improved performance of the two-dimensional array (relative to one-dimensional arrays), by comparing laser Doppler vibrometry measurements of equivalent 2-D and 1-D arrays created in silicon wafers using laser microfabrication.

10:00–10:15   Break

10:15

5aEA9. Methods of fitting a nonaudible murmur microphone for daily use and development of urethane elastmer duplex structure type nonaudible murmur microphone. Yoshitaka Nakajima and Kiyohiro Shikano (Grad. School of Information Sci., Nara Inst. of Sci. and Technol., 8916-5 Takayama, Ikoma, Nara, 630-0101 Japan)

Nonaudible murmur (NAM) microphones capturing vibrations directly from the skin without mediation by the air enable the sampling of nonaudible murmur with almost the same amplification rate as normal speech. Because they are filtered through the human body, they are robust against air-conducted noises. Soft silicone has been used, which has acoustic impedance close to human flesh, as a sound medium between the electrode of the condenser microphone and the skin by fixing it in a neckband style. We replaced the soft silicone with urethane elastomer, which has self-adhesiveness, to develop a NAM microphone that is easily wearable in everyday life without expendable supplies or supporting devices. It can be fixed to the skin, and it also decreases body movement noises. But miniaturized or sticking-type NAM microphones increased the mixture of air-conducted noises from the back of the microphone and from a gap between the skin and the rim of the microphone. To solve this problem, we created a duplex structure inside the urethane elastomer wrapping condenser microphone and outside the urethane elastomer for fixation and noise proofing to develop a miniaturized sticking-type NAM microphone that was again robust against air-conducted noise.

10:30

5aEA10. Throat-contact microphone with an elastomer-supported piezoelectric diaphragm. Daisuke Yoshioka, Hiroaki Okano, Juro Ohga (Shibaura Inst. of Technol., 3-7-5, Toyosu, Koutou-ku, Tokyo, 135-8548, Japan), Ikuo Oohira (Ashida Sound Co., Ltd., Shinagawa-ku, Tokyo, 141-0032, Japan), and Masato Miyoshi (NTT Commun. Sci. Labs., Seika-cho, "Keihanna Sci. City," Kyoto, 619-0237, Japan)

Throat-contact microphones are free from environmental acoustical noise because they pick up only local vibration of the throat. This paper proposes a new piezoelectric throat-contact microphone using an elastomer-supported diaphragm. It is smaller and lighter than conventional

electromagnetic throat-contact microphones, which have a magnet and coil. Transducer characteristic and parameters are examined in this paper to establish design procedures for such microphones. Frequency characteristics of output signals and electrical impedance for the microphones and their transducing components are presented. The piezoelectric diaphragm is supported by a few elastomer boundaries.

10:45

5aEA11. A study of bone conduction and its applications for audio signal transmission. Yusuke Watabe, Kochi Yui, Yuko Watanabe, and Hareo Hamada (Tokyo Denki Univ., 2-1200 Muzai-gakuendai, Inzai, Chiba Prefecture 270-1382, Japan)

Human can perceive a sound in two different ways: one is sound transmission through the ear canal, another is that through the skull. The bone conduction technology has been used mainly for hearing impaired listeners and as communication tools in the military organization. Recently, because of a rapid improvement of the bone conduction device, its technology has spread in the field of audio. However, the relationship between the physical characteristic and the subjective responses of bone conduction are not clearly understood. For example, the effect of attachment position and excitation force of bone conduction loudspeakers has been still under discussion. In addition, since the bone conduction units also generate the airborne sound, listeners can hear the sound composed of both the bone conduction sound and the airborne sound. In our previous work, by using of a loudness curve and a threshold of hearing, a fundamental performance of the bone conduction was discussed. In this paper, further investigation will be carried out in order to establish a transmission system of the bone conduction, and the airborne sound on the eardrum that is caused by the vibration of the bone conduction loudspeaker will be measured.

11:00

5aEA12. Acoustic communication system using bone conduction elements. Kenji Muto (Tokyo Metropolitan College of Industrial Technol., 8-52-1 Minamisenju, Arakawa-ku, Tokyo, Japan, muto@kouku-k.ac.jp), Guoyue Chen (Akita Prefectural Univ.), Kunihiko Takano (Tokyo Metropolitan College of Industrial Technol.), Kikuo Asai, and Kimio Kondo (Natl. Inst. of Multimedia Education)

This study examines an acoustic communication system constructed using a bone conduction speaker and a bone conduction microphone. This system is a headset that places a bone conduction speaker to the user's cheek, and a bone conduction microphone to the other cheek. A user can thereby hear a voice through the bone conduction speaker while hearing a voice of a loudspeaker and surrounding noise when using it for teleconferencing. It can transmit the voice of the remote loudspeaker using the external ear, but the user can hear the translated voice through bone conduction. However, the sound of the bone conduction speaker and the surrounding sound are mixed with the speaker's voice using the bone conduction microphone. The vibration level of the bone conduction microphone from the bone conduction speaker was about 0 dB, although it was the same level as the level of the speaker's voice. The vibration level from surrounding sound was about -3 dB, although it was a smaller level than the level of the speaker's voice. This device removed mixed sounds using the echo canceller [K. Muto *et al.*, Proc. Inter-noise (2004), p. 248]. The echo elimination level was about 12 dB. [This work was supported by JSPS KAKENHI 17300283.]

SATURDAY MORNING, 2 DECEMBER 2006

IAO NEEDLE/AKAKA FALLS ROOM, 8:20 TO 11:35 A.M.

Session 5aMU

Musical Acoustics: Simulation and Measurement Techniques for Musical Acoustics I

Murray D. Campbell, Cochair
Univ. of Edinburgh, School of Physics, Mayfield Rd. Edinburgh, EH9 3JZ, U.K.

Gary P. Scavone, Cochair
McGill Univ., Music Technology, Schulich School of Music, 555 Sherbrooke St., W., Montreal, Quebec, H3A 1E3, Canada

Shigeru Yoshikawa, Cochair
Kyushu Univ., Dept. of Acoustic Design, 4-9-1 Shiobaru, Minami-Ku, Fukuoka 815-8540, Japan

Invited Papers

8:20

5aMU1. Observing the effects of waveguide model elements in acoustic tube measurements. Tamara Smyth (Simon Fraser Univ., Surrey, BC, Canada) and Jonathan Abel (Universal Audio Inc., Santa Cruz, CA)

The theory of digital waveguide synthesis and its use in modeling virtual musical instruments, and in particular for cylindrical and conical bores, is well documented. Current models rely on certain approximations, however, particularly in cases where the theory provides no exact closed form solution, such as the reflection and transmission occurring at the bore's open end. In this research, we observe, from a time domain, waveguide model perspective, how the theory corresponds to actual acoustic measurements. We consider four simple acoustic tube structures, incorporating both open and closed boundary conditions for both a simple cylinder and a cylinder with a conical flare, allowing us to isolate, and observe, the filtering effects of each model component.

5a SAT. AM

5aMU2. Ingoing and outgoing waves in time-domain simulation of wind instruments. Eric Ducasse (Laboratoire de Mecanique Physique, UMR CNRS 5469, Universite Bordeaux I, 351 cours de la Liberation, F-33405 Talence Cedex, France)

An extension of a recently published work [Ducasse, J. Acoust. Soc. Am. **112**, 3031–3041 (2002)] is presented. Two-port modeling is an efficient tool for time-domain simulation of sound propagation in wind instruments. It requires the separation of ingoing and outgoing waves contributing to the acoustic field in any cross section of the air column. This separation is achieved by using a multimodal approach, where uncoupled traveling modes appear in the case of cylindrical waveguides: the propagation of the plane mode is straightforwardly modeled by a delay line and the impulsional response of any other mode is the sum of a delayed Dirac pulse and a closed-form dispersive term. Although mode coupling occurs in acoustic pipes with nonuniform cross section, the scattering matrix can be directly obtained. When a multimodal approach is not adequate, the time-domain responses of complex elements are numerically calculated by introducing either nonreflecting boundary conditions (NRBCs) or perfectly matched layers (PMLs) in a finite-element method. The examples of (*i*) the reflection matrix of a bell, as seen from the body, and (*ii*) a piece of pipe containing an obstacle (e.g., the register hole) illustrate the method.

5aMU3. Differences between cylindrical and conical brass instruments; the nonlinear propagation point of view from experiments and simulations. Joel Gilbert (Laboratoire d'Acoustique de l'Universite du Maine, CNRS, avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France, joel.gilbert@univ-lemans.fr)

The brightness of the sound generated by brass instruments is due to the essential nonlinearity of the wave propagation in the pipe. The bright instruments such as the trumpet and the trombone are different from brass instruments such as saxhorns or flugelhorns. The bright instruments have a cylindrical pipe segment just downstream of the mouthpiece. The conical bore of the saxhorns implies a faster decay of the wave, which reduces the nonlinear wave steepening. In this talk, theoretical and experimental results of brassiness associated with brass instruments having different bores, from mainly cylindrical ones to mainly conical ones, will be shown. The theoretical results come from simulations of periodic regimes using a theoretical background based on generalized Burger's equations.

5aMU4. Motion of brass players' lips during very loud playing. Seona Bromage, Murray Campbell, Samuel Stevenson (School of Phys., Univ. of Edinburgh, Edinburgh EH9 3JZ, UK, d.m.campbell@ed.ac.uk), John Chick (Univ. of Edinburgh, Edinburgh EH9 3JL, UK), and Joël Gilbert (Laboratoire d'Acoustique de l'Université du Maine, 72085 Le Mans Cedex9, France)

When a crescendo is played on a trumpet, horn, or trombone, the timbre changes markedly at the fortissimo level: there is a dramatic increase in the relative amplitudes of the high- frequency harmonics, and the resulting sound is often described as "brassy." Shock-wave generation in the air column of the instrument as a consequence of nonlinear steepening of the wavefront is the generally accepted explanation of this phenomenon. It has, however, also been suggested that saturation of the lip opening could play a role in the timbre change. This study compares the motion of the lips at lower dynamic levels and during extremely loud playing, using a high-speed digital camera and instruments with different sizes of mouthpiece. Curves illustrating the variation of the open lip area as a function of time at different dynamic levels are presented, together with waveforms and spectral analyses of both the radiated sound and the pressure in the mouthpiece. The results suggest that there is no qualitative change in the time-dependent behaviour of the opening area of the lips when the instrument enters the brassy regime.

5aMU5. Input impedance measurements of alto saxophones with a calibration error analysis. Antoine Lefebvre and Gary Scavone (Music Tech. Area, McGill Univ., Montral H3A 1E3, Canada)

Recent investigations are presented for the measurement of acoustic input impedance of wind music instruments. Initial efforts have focused on measurements of alto saxophones using a two-microphone, three-calibration tube procedure. We first present the design of the impedance probe, as well as the data sampling and analysis procedures. We then discuss the relationship between errors in the calculated calibration parameters and the resulting measured transfer functions. Finally, we compare the input impedances of Yamaha Z and Selmer Series II alto saxophones for several low register fingerings. We note several instances where the Yamaha saxophone possesses significantly stronger fundamental resonances than the Selmer saxophone.

Contributed Papers

5aMU6. Interactive program for computer-aided design of woodwind musical instruments. Héctor Alfonso Cordourier-Maruri and Felipe Orduña-Bustamante (CCADET-UNAM, Circuito Exterior CU, CP 04510, México DF, México, felipe@aleph.cinstrum.unam.mx)

Current scientific knowledge on the acoustics of woodwind instruments allows accurate physical modeling of the sound generation mechanisms, and the acoustic interaction with the air column. These physical models can be incorporated into computer software, in order to evaluate not only the tuning, voicing, and other tonal characteristics of existing instrument designs, but also to estimate the potential effects of proposed modifications, and to anticipate the characteristics of new designs. This presentation describes the development of an interactive computer program for the acoustic design, evaluation, and auralization of woodwind musical instruments. The program displays a graphical user interface that allows convenient entry of air column data, including constant and variable cross-section segments, open and closed fingerholes, and air column terminations. It also implements different types of sound excitation mechanisms, which include jet-driven (flutelike) and single reed (clarinet-like) models. The program can produce input impedance and reflection function data for further acoustic analysis, and synthetic sounds that approximate the musical tones that can be expected form the actual instru-

ment. The software has been validated by comparative analysis of calculated versus measured air columns, and simulated versus actual (recorded) flute tones.

10:35

5aMU7. Control parameters inversion using genetic algorithms applied to numerical impedance synthesis for woodwinds. Laura Perichon, Olivier Carriere, Jean-Pierre Hermand, Matthias Meyer (Environ. Hydroacoustics Lab, Optics and Acoust. Dept., ULB-CP 194/05, 50 Av. F.D. Roosevelt, B-1050 Brussels, Belgium), and Philippe Guillemain (Ctr. Natl. pour la Recherche Scientifique, 13402 Marseille cedex 20, France)

Woodwind sound synthesis in real time has been achieved using different computational techniques, especially the so-called numerical impedance model based only on acoustical variables. The impedances of the excitation mechanism and of the resonator are approximated by IIR filters and related by a nonlinear coupling. The control parameters can be reduced to a number of three including the length of the resonator, the mouth pressure, and a mask parameter. In this paper we present preliminary inversion results for the latter two parameters in the case of a clarinet. The inversion algorithm is based on differential evolution metaheuristics with a number of prior conditions on the correlated parameters. A cost function based on time-domain correlation has been implemented, allowing us to obtain better results in terms of signal matching than with conventional descriptors extraction. The technique has proved its effectiveness on the sustained part of synthesized sounds and a regularization method is presented to allow us to invert attack and decay parts as well as real analogically recorded signals. Results will be of main interest for tuning control parameters in real-time model-based synthesis.

10:50

5aMU8. Virtual wind instruments based on pulse forming synthesis. Michael Oehler (Inst. for Appl. Musicology and Psych., Duerener Strasse 98, 50931 Cologne, Germany, oehler@iamp.info) and Christoph Reuter (Univ. of Cologne, 50923 Cologne, Germany)

A new synthesis method for wind instruments is presented, the digital pulse forming. The core of that principle is that every wind instrument sound can basically be put down to its excitation impulses, which independently of the fundamental always behave according to the same principles. By controlling the pulse width and shape it is possible to generate all sound nuances that can be produced on a real wind instrument. Based on the 1975 found principles of generating wind-instrument-like spectra with typical stable formant areas and spectral gaps evoked by the excitation pulses of double-reeds or lips [J. P. Fricke, Fortsch. Akust. **4**(DAGA75), 407–411 (1975)], a virtual wind instrument, the Digital Variophon, is developed. The resulting software-based version of the original Variophon, the first analog wind synthesizer using the pulse forming principles, is a further step towards the intended scientific experiment system for analyzing and synthesizing (wind) instrument sounds. Trumpet, bassoon, and oboe modules are already digitally implemented in a flexible, realtime sound generating and analyzing framework. Acoustic analyses and hearing experiments show extensive concordance between the original instruments and the Digital Variophon. [Work supported by DFG.]

11:05

5aMU9. Flow regimes at the output of oboe double-reeds. Christophe Vergez (LMA, CNRS UPR-7051, 31 chemin Joseph-Aiguier, 13402 Marseille, cedex 20, France) and Reneé Caussé (IRCAM)

Measurements of the quasistatic nonlinear characteristic curve for double reeds show systematic deviations from an elementary model of a reed instrument based on a spring model associated to a Bernoulli flow input into the reed. These deviations were previously explained by the pressure recovered by the flow as it expands in the conically divergent output duct of the reed. We will show measurements of the pressure recovery inside the reed as well as detailed flow profiles that can explain the different regimes observed for the pressure recovery (laminar and turbulent). An extension of these measurements to oscillating regimes allows us to observe some qualitative differences in the flow, suggesting an important influence of the dynamic terms in the Navier-Stokes equation in these kinds of regimes.

11:20

5aMU10. Wolf notes in brass wind instruments. Thomas R. Moore and Isaac E. W. Codrey (Dept. of Phys., Rollins College, Winter Park, FL 32789)

It has been recently reported that experiments using artificial lips have shown that damping the bell vibrations of a trumpet can result in changes in the sound spectrum. New data now suggest that these effects are enhanced when the fundamental frequency of the note is near a resonance frequency of the bell vibration. This effect results in a situation that is analogous to the wolf note commonly found on stringed instruments. Although the physics of the process is still poorly understood, the effect appears not to be related to a change in impedance of the air column.

5a SAT. AM

Session 5aNS

Noise and Physical Acoustics: Prediction and Propagation of Outdoor Noise I

D. Keith Wilson, Cochair

U.S. Army Cold Regions Research Lab., Engineering Research and Development Ctr., 72 Lyme Rd., Hanover, NH 03755-1290

Kohei Yamamoto, Cochair

Kobayashi Inst. of Physical Research, 3-20-41 Higashi-Motomachi, Kokubunnji-shi, Tokyo 185-0022, Japan

Chair's Introduction—7:30

Invited Papers

7:35

5aNS1. Sound field modeling in a street canyon with a diffusion equation. Judicaël Picaut, Stéphane Colle, and Michel Bérengier (LCPC, Section Acoustique Routière et Urbaine, Rte. de Bouaye, BP 4129, 44341 Bouguenais Cedex, France)

The transport theory of sound particles is applied to the sound field modeling in an empty street canyon with partially diffusely reflecting facades. A diffusion equation is then derived to predict the sound field distribution and the sound decay in the street. The main parameter, namely the diffusion coefficient, is a function of the street width and the acoustic reflection laws of both building facades. For a single rectangular street, analytical solutions can be found. For more complex urban spaces, a finite-element based approach is proposed. Several numerical examples are given, like the sound propagation in a street with a nonuniform cros section, and in street intersections, for point and line sources. In comparison with ray-tracing-based models, the diffusion model requires less computation time and could be applied to the calculation of a sound map for large urban areas. However, at the present time, the diffusion coefficient can be calculated only for simple building reflection laws, like the Lambert's one.

7:55

5aNS2. Prediction of road traffic noise using two-dimensional numerical analysis. Shinichi Sakamoto (Inst. of Industrial Sci., The Univ. of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, sakamo@iis.u-tokyo.ac.jp), Akinori Fukushima (NEWS Environ. Design, Inc., Hyougo, Kobe 652-0802), Tomonao Okubo, and Kohei Yamamoto (Kobayasi Inst. of Physical Res., Kokubunji, Tokyo 185-0022, Japan)

For acoustically complicated road structures such as semi-underground roads and special areas in which a viaduct road and a flat road with noise barriers exist together, prediction of road traffic noise is complicated because of multiple reflections and diffractions that occur inside the road structures. For such road structures, an energy-based engineering model cannot be applied and noise propagation should be addressed through introduction of wave theory. When the road structures have almost identical cross-sectional shape along the road, two-dimensional (2-D) wave-based numerical analyses are applicable. In the prediction model of road traffic noise, the ASJ-RTN Model 2003, published by the Acoustical Society of Japan (ASJ), the application of 2D wave-based numerical analysis was introduced as a prediction method for such complicated road structures. Comparisons between calculations by BEM and FDM and field measurements and experiments for three actual road structures were conducted. Consequently, calculation results agreed well with measured ones. Therefore, the validity of the calculation methods was confirmed. This research was discussed in the Research Committee of Road Traffic Noise in ASJ. Measurement data were provided by Nippon Expressway Company and the Nagoya Expressway Public Corporation.

8:15

5aNS3. Modeling outdoor sound propagation in mountainous areas. Dick Botteldooren, Timothy Van Renterghem, and Bram de Greve (Acoust. Group, Dept. of Information Technol., Ghent Univ., St. Pietersnieuwstraat 41, 9000 Gent, Belgium, dick.botteldooren@ugent.be)

Detailed modeling of outdoor sound propagation in mountainous area imposes special requirements. The strong gradients in ground surface create propagation conditions that may lead to noise levels much higher than expected at some distance from the source partly due to a focusing effect. Meteorological conditions are particular: main winds following the valley, strong temperature gradients often including stable inversion, slope winds, etc. Because of these particular conditions, efficient modeling of outdoor sound propagation in mountainous area requires adaptations to be made to general purpose models. In this paper we will present possibilities for making time-domain models terrain following and consider options to decrease the memory requirements. For the higher frequency range and distances one is generally interested in, these adaptations are insufficient. Hence, a Green's function parabolic equation model is added to the toolbox. To tackle the problem in three dimensions, a number of two-dimensional slices connecting the receiver point to all (segments of) sources considered is made. Propagation results for each slice are added. Numerical simulations are compared to field measurements made close to an Alpine highway.

5aNS4. Measurements of acoustic impedance and their application to sound propagation calculations and audible simulations.
Teruo Iwase (Faculty of Eng., Niigata Univ., 8050 2nocho, Niigata City, 950-2181, Japan), Yu Murotuka (Niigata Univ., Niigata City, 950-2181, Japan), and Koichi Yoshihisa ((Meijo Univ., Tempaku-ku, Nagoya, 468-8502, Japan)

Acoustic impedance is an important value that determines the boundary condition of each sound field, but collections of actual values are not sufficient for evaluation of many sound fields. First, measurements using a particle velocity sensor for acoustic impedance were tested on various fields. Such measurement results were applied to calculations of sound propagation. Frequency characteristics of sound propagation were obtained along such surfaces as fields of lawns and snow, and areas paved by porous asphalt for drainage. Those characteristics showed fair correspondence with inspected field measurement results. Then, fine calculations in the frequency domain were converted to the impulse response for each sound field model. Convolution operations based on the impulse response and on voice, music, and some noise sources readily produced an ideal sound field to the audible sound file. Furthermore, simulations of the noise, including noise reduction effects from a car running through a drainage pavement area, were executed as advanced applications. Our measurement method of acoustic impedance is inferred to be useful. Calculated sound propagation characteristics based on the measured acoustic impedance correspond to actual propagation characteristics. Audible simulation will be effective as a future means of sound-field evaluation. [Research supported by KAKENHI.]

5aNS5. Quiet heights and pulse tails: The modal structure of near ground to near ground sound propagation in the nocturnal duct. Roger Waxler (NCPA, Univ. of Mississippi, University, MS 38677, rwax@olemiss.edu)

On clear nights over land a temperature inversion forms, creating an acoustic duct in the first few hundred meters of the atmosphere. Assuming a flat, locally reacting ground surface and a stratified atmosphere, the sound field can be expressed as a superposition of independently propagating vertical modes [J. Acoust. Soc. Am. **112** (6), 2540–2552 (2002) and **115** (4), 1437–1448 (2004)]. The modes account for refraction due to sound-speed gradients as well as for the compliance and resistance of the ground. For typical nocturnal ducts and ground surfaces one finds two types of modes: a surface mode and a set of higher modes. The surface mode is more slowly propagating and more rapidly attenuated than the higher modes. The higher modes' amplitudes all have a sharp minimum several meters above the ground. Two consequences of this modal structure have been observed: At long ranges, narrow-band signals have a range-independent quiet height at which sound levels are reduced by 10 to 15 dB from the levels on the ground [J. Acoust. Soc. Am. **119** (1) 86–95 (2006)] while impulsive signals develop a bandlimited low-frequency tail [J. Acoust. Soc. Am. **88** (1), 455–461 (1990)].

5aNS6. Influence of short-term variations of meteorological parameters on sound propagation outdoors. Philippe Blanc-Benon, Benjamin Cotte (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France), Benoit Gauvreau, and Michel Berengier (Laboratoire Central des Ponts et Chausses, BP 4129, 44341 Bouguenais Cedex, France)

Predicting long-range sound propagation over a nonurban site with complex propagation media requires the knowledge of micrometeorological fields in the lower part of the atmospheric boundary layer. In the framework of road traffic noise characterization there is a need for reliable sound pressure level predictions for specific propagation conditions that must be representative of time and space (small scale and site effects) characteristics of the acoustic situation. Outdoor measurements involving roughly 100 meteorological and acoustic sensors have been carried out during 3 months in 2005 in the southwest of France. This large database enables us to study the variability of meteorological parameters (wind speed, temperature, etc.) on different time scales. Scaling parameters in the atmospheric surface layer (such as the friction velocity and the Monin-Obukhov length) are estimated every 15 min, which allows looking at relatively short-term variations. Some longer-term effects, from days to months, are studied too. These include occurrences of meteorological configurations and variation of ground parameters. The effect of meteorological variability on sound propagation is studied using parabolic equation simulations. Comparisons between numerical predictions and experimental data will be discussed.

5aNS7. Effect of atmospheric absorption on aircraft noise propagation in various world regions during the year. Yasuaki Okada, Koichi Yoshihisa (Meijo Univ., 1-501 Shiogamaguti, Tempaku-ku, Nagoya, 468-8502, Japan), and Teruo Iwase (Niigata Univ., Niigata, 950-2102, Japan)

The attenuation of sound due to atmospheric absorption depends strongly on environmental temperature and humidity. For that reason, it varies in response to changing meteorological conditions on a variety of time scales. To perform long-term prediction of outdoor sound propagation more accurately, it is necessary to examine the fluctuation of the attenuation coefficients for atmospheric absorption during the year. Large variations occur according to the time and place in actual meteorological conditions, on which the attenuation coefficients depend strongly. The attenuation coefficients for atmospheric absorption were calculated from hourly meteorological data obtained at ten international airports throughout the world during a year using the calculation method described in ISO 9613-1. Calculated results show that the variability characteristics of the attenuation coefficients during the year depend strongly upon both the sonic frequency and the place. The effects of atmospheric absorption on aircraft noise propagation were examined using the calculated attenuation coefficients. Results showed that the A-weighted sound pressure levels and sound spectra of aircraft noise vary strongly because of the changes in atmospheric absorption, which depend on actual meteorological conditions of various world regions.

5a SAT. AM

5aNS8. Acoustic travel-time tomography of temperature and wind velocity fields in the atmosphere. Vladimir E. Ostashev (NOAA, Earth System Res. Lab., Boulder, CO, and Dept. of Phys., New Mexico State Univ., Las Cruces, NM), Sergey N. Vecherin (New Mexico State Univ., Las Cruces, NM), D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH), Astrid Ziemann, Manuela Barth, and Klaus Arnold (Inst. of Meteorol., Univ. of Leipzig, Germany)

Acoustic tomography of the atmosphere allows one to reconstruct (estimate) temperature and wind velocity fields in the atmosphere and to monitor the evolution of these turbulent fields in time. In the present paper, we report on a progress in construction of a state-of-the-art array for acoustic tomography of the atmosphere, which will allow us to measure travel times of sound propagation between different pairs of sources and receivers within a few meters above the ground. Using these travel times, the turbulent fields will then be reconstructed using different inversion algorithms. The array is being built at the Boulder Atmospheric Observatory, CO, in a collaborative effort between several organizations in the U.S. Furthermore, we discuss several inverse algorithms for estimation of the turbulent fields, including recently developed time-dependent stochastic inversion. Finally, some of these algorithms were used to reconstruct temperature and wind velocity fields in indoor and outdoor tomography experiments carried out by scientists from the Institute of Meteorology, University of Leipzig, Germany. Examples of the reconstructed fields are given. [Work supported by ARO, Grants DAAD19-03-1-0104 and W911NF-06-1-0007.]

10:15−10:30   Break

Contributed Papers

10:30

5aNS9. A comprehensive noise study for the city of Lincoln, Nebraska. Dominique Cheenne, Connie Lee, Michael Cappiello, Sam Lalk, Coleman Martin, and Philip Muzzy (Dept. of Audio Arts & Acoust., Columbia College Chicago, Chicago, IL 60610)

A comprehensive noise study to define the soundscape of the city of Lincoln, NE, has been undertaken by the students enrolled in the Acoustics Program at Columbia College, Chicago. The study includes (1) a global review of best practices in the field of environmental noise control using a perspective emphasizing close attention to effective and proven solutions that can be implemented via policies and effective legislation, (2) the creation of a large-scale outdoor noise propagation computer model that incorporates all primary traffic arterial and readily identifiable noise sources, (3) a comprehensive large-scale testing of noise levels using specially modified noise dosimeters allowing for determination of Leq, and Ldn over 24-h intervals, and (4) a large scale attitudinal survey of the population using a web-based survey engine that allows for the geographical location of the respondents. The results of the survey are used to refine the density of the test locations and the scale resolution of the software model, as well as to provide city and county officials with a targeted course of action pertaining to regulations and noise abatement policies. The project is expected to last 3 years with up to 1200 test locations being analyzed.

10:45

5aNS10. Effect of parameters on modeling low-frequency propagation of petrochemical plant noise over water. Frank Brittain (Bechtel Corp., 2255 Peavine Valley Rd., Reno, NV 89523, fhbritta@bechtel.com) and Xiao Di (Univ. of Mississippi, University, MS 38677)

Modeling noise is vital to designing petrochemical plants to meet community noise limits. Models are used to specify equipment noise, choose add-on controls, and confirm limits will be met. For propagation over land, there are substantial variations of actual levels from the long-term average computed using ISO 9613-2. Commercial modeling software based on ISO 9613-2 excludes propagation over water, where levels are often higher and more variable than at the same distance over land. Over water, higher levels and variability arise primarily from wind, thermal inversions, and lack of attenuations from ground clutter. Often a petrochemical plant and receivers are separated by a large body of water, which reduces modeling accuracy. To better understand propagation over water, full-wave Green's Function-Parabolic Equation (GF-PE) software was used to predict levels between 50 m and 10 km for various lapse rates (temperature increase with elevation) as well as inversion, source, and receiver heights. Plots of attenuation with distance and contours in a vertical plane are presented for 63 Hz. Rates of attenuation considerably less than −6 dB

per distance doubling for spherical spreading were found; these and implications for design are discussed. The potential application of GF-PE to verify algorithms for outdoor propagation is indicated.

11:00

5aNS11. Bauxite refinery community noise control and prediction. Jose Augusto Nepomuceno (Acustica & Sonica, R. Fradique Coutinho, 955, sala 01, Sao Paulo, 05416-011 SP, Brasil, janepomuceno@yahoo.com) and Aljan Machado (Alcoa Brasil)

Alumar in the city of Sao Luis in Brazil is one of the largest bauxite processing plants in the world including refinery and reduction plants. The Alumar refinery will expand from approximately 1.4 million mtpy to approximately 3.5 million mtpy and the expansion is expected to be completed in the first half of 2008. Since the year 2000 the authors worked in collaborative way with the in-plant and community noise control and prediction program for the expansion project. The program included noise measurement in the existing plant, development of the Bauxite Refinery Noise Control Manual, several noise control guidelines, and noise modeling to predict community noise impact. The plant is surrounded by different land uses as residential, industrial, rural, and natural permanent protection. Using computer modeling it was studied how the refinery expansion could impact the community areas and what will be the critical equipment requiring special noise limits. The validation process predicted measured results in the range of +2.2 dB(A). The paper presents the noise program results obtained by the year 2006.

11:15

5aNS12. Statistical models for fading and coherence of sound in urban environments. D. Keith Wilson (U.S. Army Engineer Res. and Development. Ctr., 72 Lyme Rd., Hanover, NH 03755, d.keith.wilson@erdc.usace.army.mil), Rafael Bey-Hernandez (U.S. Army Engineer Res. and Dev. Ctr., Hanover, NH 03755), and Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305)

Urban environments typically produce strong, multipath scattering. It is often desirable to characterize the effects of the scattering with statistical models for signal fading and coherence. This paper discusses initial efforts to develop such models. Random configurations of buildings, with varying sizes and number densities, are synthesized. Sound waves are then propagated through the random configurations. The first and second moments of the sound field are calculated by averaging results from the random realizations. Most of the propagation calculations are done by a ray-tracing technique, which is very fast but does not include diffraction. Due to multiple scattering, cases where a receiver is in a full acoustic shadow are infrequent. We also compare the ray tracing to finite-difference, time-domain calculations, which provide a full-wave solution that includes diffraction as well as multiple scattering. Initial results show

that the wave extinction and coherence diminish exponentially with distance, as is the case for turbulent scattering. The extinction and coherence decay rate increases with the number of buildings and size, but appears to be independent of frequency. We compare the signal probability density functions to statistical fading models developed for radio-wave scattering in urban environments.

11:30

5aNS13. A review on barrier diffraction theories and related insertion loss comparisons in highway noise models. Ning Shu (AZTEC Eng., 4561 East McDowell Rd., Phoenix, AZ 85008), Louis F. Cohn, Roswell A. Harris, and Teak K. Kim (Univ. of Louisville, Louisville, KY 40292)

This paper presents a review on diffraction theories for highway noise barriers. Insertion loss based on different diffraction theories was investigated with two highway noise models, STAMINA and TNM 2.5. STAMINA implements the Kurze and Anderson empirical diffraction formula while TNM 2.5 deploys a simplified MacDonald analytical model. Quantitative analysis indicates that from the perspective of diffraction theory, the average difference of insertion loss between TNM and STAMINA for a point source is about 3 dB(A). To improve the prediction accuracy of TNM, it is recommended that the MacDonald solution, but with real and image sources, should be deployed.

11:45

5aNS14. An acoustic technique for monitoring off-highway vehicle (OHV) use with sound level meter data: Camparison with Trailmaster records. Tina Yack, Ann Bowles (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 95616, tyack@hswri.org), Keith Slauson, William Zielinski (USDA Forest Service, Arcata, CA 95521), and Kenneth Plotkin (Wyle Labs., Arlington, VA 22202)

Off-highway vehicle (OHV) use is of growing interest to land managers. Trail counters are typically used to monitor OHV activity, but accuracy of this technique remains relatively untested. As part of an ongoing study comparing the distribution of American martens (*Martes americana*) with OHV use in the Lake Tahoe Basin, verifiable OHV detections were collected using two techniques, (1) photographs from Trailmaster 1500 camera systems (TMs) placed across roads/trails and (2) continuous 2-s, A-weighted equivalent-continuous sound levels (LAeq2s) collected using Larson-Davis 720 Sound Level Meters (SLMs) placed either near roads/trails or paired with TM animal detection stations. Events were identified in the SLM data by comparing time-history profiles collected in the field with profiles of known snowmobile passes. Errors in SLM event counts were assessed using independent observer logs and compared with the TM records. SLMs were consistently more reliable detectors of snowmobile use. During the Winter 2004 season, TM records were reliably matched to SLM logs only 48% of the time. TMs consistently missed passes and overestimated events compared to SLMs. In conclusion, SLMs proved to be more sensitive and reliable detectors of OHV activity. These results have implications in both wildlife research and land management applications.

SATURDAY MORNING, 2 DECEMBER 2006

WAIANAE ROOM, 7:30 A.M. TO 12:00 NOON

Session 5aPA

Physical Acoustics: Thermoacoustics

Tetsushi Biwa, Cochair
Tohoku Univ., Dept. of Mechanical Systems and Design, Aoba-ku, Sendai 980-8579, Japan

Steven L. Garrett, Cochair
Pennsylvania State Univ., Graduate Program in Acoustics, State College, PA 16804-0030

Contributed Papers

7:30

5aPA1. Amplification of acoustic intensity of a pulse wave propagating through a tube having a temperature gradient. Tetsushi Biwa (Dept. of Mech. Syst. Design, Tohoku Univ., Aoba-ku, Sendai, 980-8579, Japan), Yusuke Tashiro, Masatoshi Shinmei (Nagoya Univ., Chikusa-ku, Nagoya, 464-8603, Japan), and Taichi Yazaki (Aichi Univ. of Education, Igaya-cho, Kariya, Japan)

Ceperley has theoretically pointed out that when acoustic traveling waves pass through a long duct with a positive temperature gradient, the acoustic intensity is amplified as a result of the thermodynamic cycles that the gas undergoes. Even though he only observed damping of the acoustic intensity in the experiment, his idea serves as a starting point of new acoustic devices known as thermoacoustic Stirling heat engines. For a further progress of the thermoacoustic technology, experimental verification of his idea is of great importance. In this paper, we report on the thermoacoustic amplification of acoustic intensity using pulse waves traveling in a long tube equipped with a differentially heated regenerator. A reservoir filled with pressurized air and a solenoid valve were used to generate the acoustic pulse running in one direction in the tube with length 60 m. A series of pressure transducers was used to determine the axial distribution of the acoustic intensity. Use of a pulse wave enabled us to see

the frequency dependence of the amplification rate of acoustic intensity very easily. We found that the amplification rate was essentially determined by the ratio of thermal boundary layer thickness to the pore radius of the regenerator.

7:45

5aPA2. Pattern formation in thermoacoustic resonators. German J. de Valcarcel (Dept. de Optica, Universitat de Valencia, Dr. Moliner 50, 46100 Burjassot, Spain), Victor J. Sanchez-Morcillo, and Isabel Perez-Arjona (Universidad Politecnica de Valencia, 46730 Grao de Gandia, Spain)

The phenomenon of pattern formation is ubiquitous in the dynamics of extended nonlinear systems. Patterns arise from the interplay of external forcing, dissipation, nonlinearity, and spatial coupling. These factors are also present in many acoustic systems. However, under usual conditions where the elastic nonlinearity dominates, the development of shock fronts in nonlinear regime strongly limits the possibility of a simple analysis. Viscous media, where thermal nonlinearity dominates over the elastic one and wave distortion is inhibited at short distances, are good candidates for acoustic systems presenting patterns with similar characteristics as those found in other fields, such as hydrodynamics or optics. A model describing

the evolution of pressure and temperature variations in a resonator with plane walls containing a viscous medium is presented. The large aspect ratio of the resonator allows the transverse spatial coupling, mediated by sound diffraction and temperature diffusion. The homogeneous solutions show bistability at some values of the parameters, in agreement with experimental results reported in the bibliography. We show by means of linear stability analysis that the homogeneous solutions develop spatial instabilities leading to the spontaneous generation of spatial patterns. The analytical predictions are confirmed by a numerical analysis. [Support from Spanish MEC, Project FIS2005-07931-C03-02/03, is acknowledged.]

8:00

5aPA3. Anharmonic acoustic resonators in miniature thermoacoustic cooler. H. El-Gendy, L. Lyard, and O. G. Symko (Dept. of Phys., Univ. of Utah)

In optimizing miniature thermoacoustic coolers for high cooling power density, the acoustic resonator has to be able to sustain very high amplitude acoustic oscillation at the resonant frequency of the driver in order to pump heat up a temperature gradient. Nonlinear effects in cylindrical-shaped resonators lead to acoustic saturation; acoustic energy is dissipated in higher harmonics, thus limiting the intensity of the fundamental mode. To overcome this limitation, noncylindrical resonators were studied here. Conical, exponential, and halfcosine shapes were used. They produce high-amplitude sound waves with essentially no shock waves. Such resonators were chosen here to operate at around 4 kHz, the resonant frequency of the piezoelectric driver. They are easily adapted to the miniature refrigerator. Moreover, they provide an impedance match, which is shape dependent, between the driver and the resonator. Tests at sound intensity levels up to 170 dB show essentially no distortion of the acoustic waveform. Results of their performance will be discussed. [Work supported by ONR and the State of Utah.]

8:15

5aPA4. A study for applying lower temperature heat source to loop-tube thermoacoustic cooling system. Yousuke Imamura, Shin-ichi Sakamoto, and Yoshiaki Watanabe (Faculty of Eng., Doshisha Univ., 610-0321 1-3 Tataramiyakotani Kyotanabe Kyoto, Japan)

To realize the generation of self-sustained sound in lower temperature heat source, a 5-μm in thickness rubber membrane was applied in the loop-tube thermoacoustic cooling system. It is known that the existence of dc flow acts as the negative effect for the loop-tube. The membrane reduces a loss of thermal energy supplied to the prime mover by suppressing the dc flow. So it is considered that lower temperature difference between both the edges of the stack can be realized by using presented system. The locating position of the membrane was moved, and then the temperature differences of the stack were observed. The experiments were carried out on the condition that the thermal energy supplied to the prime mover was constant. It was found that the temperature difference is affected by the position of the membrane. When the membrane was set at the approximate antinode of pressure, the lowest temperature difference was realized. In this condition, the temperature difference with the membrane was 430 K, which was 160 K lower than without the membrane. These results suggest that a lower temperature heat source can be available by setting the membrane on the loop-tube thermoacoustic cooling system.

8:30

5aPA5. Marginal instability of thermoacoustic oscillations of a gas in a tube. Nobumasa Sugimoto and Masatomi Yoshida (Dept. of Mech. Sci., Grad. School of Eng. Sci., Osaka Univ., Toyonaka, Osaka 560-8531, Japan, sugimoto@me.es.osaka-u.ac.jp)

This paper examines a marginal condition of instability of thermoacoustic oscillations of a gas in a tube with one end open and the other closed by a flat wall, subjected to a smooth temperature distribution axially. Assuming a boundary layer is thin compared with the tube radius, the linear theory is developed in the framework of the first-order theory in its

thickness. An idea of the method of renormalization enables us to obtain analytical solutions and to derive the marginal condition when the temperature distribution is parabolic. Solving the condition numerically, the marginal curve for the temperature ratio is displayed graphically against the tube radius relative to the boundary-layer thickness. It is found that the temperature ratio has a minimum and that the curve has two branches with respect to the minimum. While the left branch for viscous mode extends to infinity, the right branch, close to the curve for neutral oscillations, asymptotes a certain temperature ratio as the tube radius increases. Such results should be compared with the ones obtained by Rott in the case of a step distribution. Spatial mode of oscillations is also displayed in the marginal state and some discussions on the energy balance are included.

8:45

5aPA6. Direct observation of thermoacoustic energy conversion. Yusuke Tashiro (Dept. of Crystalline Mater. Sci., Nagoya Univ., Furo-cho Nagoya 464-8603, Japan, tashiro@mizu.xtal.nagoya-u.ac.jp), Tetsushi Biwa (Tohoku Univ., Aoba-ku Aramaki Sendai 980-8579, Japan), and Taichi Yazaki (Aichi Univ. of Education, Kariya 448-8542, Japan)

A gas column starts to oscillate when the externally imposed temperature gradient exceeds a critical value. Also, when an acoustic wave propagates through a differentially heated regenerator, the acoustic intensity is thermally amplified. Such thermoacoustic phenomena attract considerable interest not only because it becomes possible to develop pistonless Stirling heat engines, but because an alternative approach to understand a heat engine has been proposed based on energy flows. In this work, we report the direct observation of the thermoacoustic energy conversion in a tube filled with atmospheric air to test the validity of the proposal. A gas column was partly heated by an electrical heater and driven by an oscillating piston at 1.00 Hz to assure good thermal contact with the surrounding tube wall. We simultaneously measured pressure and velocity oscillations to determine the acoustic intensity. As a result, we found the amplification of acoustic intensity in the region with a positive temperature gradient. This supports the proposed approach. Further test was made by comparing the acoustic power produced per unit length and the thermodynamic cycles of the gas, the latter of which was also experimentally determined from measurements of temperature and pressure.

9:00

5aPA7. Study on a thermoacoustic cooling system to drive the fundamental resonance frequency by connecting a triggered tube. Hideo Yoshida, Shin-ichi Sakamoto, and Yoshiaki Watanabe (Ultrasonic Labs., Doshisha Univ., 1-3 Miyakodani Tatara, Kyotanabe City, Kyoto 610-0321 Japan, dtf0166@mail4.doshisha.ac.jp)

A thermoacoustic cooling system to drive a looped-tube with the fundamental resonance frequency by connecting a triggered tube is proposed. The heat pump is located in the triggered-tube. The existence of dc flows, which is caused by harmonics in the looped tube, has been pointed out as negative effects. The presented thermoacoustic cooling system, however, connected the triggered-tube is expected to suppress the harmonics and to improve cooling effect, so that the length of the triggered tube is designed to have the fundamental resonance frequency. It is well known that the phase relation between sound pressure and particle velocity is important, and approximately in-phase relation is the best position for cooling effect. In the presented system, the sound energy is supplied from the looped tube. It is important to know the best connecting position where the phase relation is approximately in phase. The sound pressures were observed, then the phase relations were calculated by two-sensor power method. By using results, the best position of the triggered tube was decided. It is confirmed that the presented system can realize the suppression of harmonics and the decrease of outlet temperature. It is also confirmed that the decrease of temperature has the higher effects compared with the looped tube in the same environmental conditions.

5aPA8. Determination of complex propagation constant, acoustic intensity, and acoustic power in an arbitrarily terminated pipe using laser Doppler anemometry. Ki Won Jung and Anthony A. Atchley (Grad. Program in Acoust., Penn State Univ., University Park, PA 16802, kuj102@psu.edu)

Microphones are widely used to determine acoustical parameters, such as the complex propagation constant and acoustic intensity and power, of sound fields. It is less common to use measurements of acoustic particle velocity for the same purpose. Precise measurement of acoustic particle velocity can be achieved using laser Doppler anemometry (LDA). Although this measurement technique requires optical access to the region of interest and the use of seeding particles, in some circumstances it can be less invasive and offer greater flexibility than measuring pressure fields with microphones. The focus of this research is to use LDA to determine the complex propagation constant, acoustic intensity, and power inside a constant cross-section circular pipe. The measured particle velocities are fit to a counterpropagating plane-wave model to determine the complex amplitudes and propagation constant. Other acoustic quantities, such as the radial-dependent acoustic intensity and the cross-sectional-averaged acoustic power, can then be calculated. The complex propagation constant is compared with the theoretical value based on thermoviscous boundary layer theory. The acoustic intensity and power are compared with numerical solutions to Rott's wave equation. [Work supported by the Penn State Graduate Program in Acoustics.]

9:30

5aPA9. Spectral gaps and discrete transmission in slender tubes. Manvir S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla, Pue. 72570, Mexico), A. Akjouj, B. Djafari-Rouhani, L. Dobrzynski, and J. Vasseur (Univ. of Sci. and Technol. of Lille-I, France)

Extensive band structure and transmission spectra for the longitudinal (acoustic) wave propagation in a system made up of N' dangling side branches (DSBs) periodically grafted at each of the N equidistant sites on a slender tube are reported. A periodic pattern of large stop bands is obtained for the airy DSB on a slender tube. The emphasis is laid on the interesting result of huge gaps and discrete transmission spectrum due only to the DSB grafted at a *single* site ($N=1$) on the slender tube. Designing the system with open tubes allows achievement of lowest gap below a threshold frequency and extending up to zero—thereby providing an entirely discrete band structure and transmission spectrum. This should have important consequences for the suppression of low-frequency noise and for designing filters and transducers.

9:45

5aPA10. The effects of turbulence on an undular bore. Pablo Luis Rendon (CCADET, Universidad Nacional Autonoma de Mexico, Ciudad Universitaria, Mexico D.F. 04510, Mexico, rendon@aleph.cinstrum.unam.mx)

The Korteweg-de Vries-Burgers (KVB) equation is the simplest equation to incorporate simultaneously the effects of nonlinearity, damping, and dispersion, and is of the form $\eta_t + \eta, \eta_x + \eta_{xxx} = \delta, \eta_{xx}$. It has been found that the steady solution to this equation exhibits all the characteristics of the undular bore, with $\eta(x,t)$ the surface profile. An examination of the effects of turbulence on undular bores, and of whether this turbulence can effectively alter the overall shape of the bore, is conducted through numerical simulation of the KVB equation. [The author thanks the Universidad Nacional Autonoma de Mexico for funding this project through a PAPIIT (IN116205) grant.]

10:00–10:15 Break

10:15

5aPA11. Resonant aero-acoustic excitation of cavity depth modes. Boris M. Efimtsov, Alexey Yu. Golubev (TsAGI, 17 Radio St., 107005 Moscow, Russia), and Anders O. Andersson (Boeing MC 67-ML, Seattle, WA 98118)

Results are generalized of parametric experimental investigations of resonant aero-acoustic excitation of the depth modes of a flow-grazed cavity at low-subsonic (U50 m/s) and transonic flows (0.7M1.045). Special attention is given to the fundamental mechanisms responsible for this phenomenon. The conditions for these mechanisms to manifest themselves and interact are predicted. Observations are made of the relations between the characteristic flow and cavity parameters at which the maximum resonant aero-acoustic response of the cavity depth mode takes place. These relations take into account the distinction between the flow velocity and the characteristic spreading velocity of the disturbances downstream of the wall discontinuity as well as the distinction between the cavity depth and the characteristic linear scale in the Helmholtz number connected with the edge effect for sound waves in the vicinity of the cavity gap. Experimental data illustrating the reliability of these relations are presented. The extent of the Strouhal-number region where the cavity mode excitation can be treated as an aero-acoustic resonance is determined.

10:30

5aPA12. Single transducer parametric sound source. Bastian Epp and Volker Mellert (Univ. of Oldenburg, Faculty V, Inst. of Phys., D-26111 Oldenburg, Germany)

The idea of using the nonlinear effects of the propagation of sound waves with finite amplitude to build up highly directed sound sources is well known. In underwater acoustics, such parametric arrays are widely used. There are also attempts to build up parametric arrays in air to generate highly directed air-borne sound. To be able to use nonlinear effects in air as a carrier medium, one needs high amplitudes of the propagating sound wave. Consequently, a superposition of single sound sources (arrays) is often reported in the literature to increase the size, and so the directivity, of the sound source as well as the acheived sound pressure level. A major disadvantage of using arrays is the problem of slightly varying resonance frequencies and phase characteristics of the single source components as well as a complicated superposition of single directivity patterns. It is difficult to investigate basic properties of wave propagation and the nature of distortion in such a way. In this work, a single, high-power ultrasound transducer was used to radiate sound waves with finite amplitude, able to generate extraneous frequency components in air. Using only a single transducer, it was possible to investigate basic properties of nonlinear wave propagation in air.

10:45

5aPA13. Scale model experiments over curved rough surfaces. James Chambers and Andrew Whelan (Natl. Ctr. for Physical Acoust., The Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

Previous research has indicated that the propagation of sound over a curved surface can be drastically altered by the presence of continuous small scale roughness that is smaller than the acoustic wavelength. This effect can alter sound levels in the shadow zone from +6 to −20 dB relative to levels past a similar smooth surface depending on the frequency of interest. These findings were originally observed for a single radius of curvature and a limited range of roughness sizes. The work has been revisited for a larger range of radii of curvature and roughness sizes with comparable results. The results for this problem as well as the related one of propagation in an upwardly refracting medium will be discussed. [Work supported by the Army Research Laboratory.]

5a SAT. AM

5aPA14. Low-frequency pulse propagation over irregular terrain.
Xiao Di and Kenneth E. Gilbert (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

The shape of a low-frequency pulse that propagates to the far side of a hill depends primarily on two physical mechanisms: a diffracted wave and a surface wave. Both of these mechanisms depend on frequency, so that the pulse shape and frequency spectrum of the received pulse differ significantly from those near the source. As is well known, diffraction acts like a low-pass filter. The surface wave has a more complicated frequency dependence that depends on the terrain, the frequency dependence of the ground impedance, and the near-ground sound speed profile. For a given range, these parameters determine a frequency band that contains most of the surface wave energy. The surface wave thus functions as a low-frequency bandpass filter. The combined effects of diffraction and surface wave propagation on pulse propagation over a hill are studied using a parabolic equation model and a piecewise linear mapping to flatten the irregular terrain. Numerical calculations are presented for selected terrain features, ground impedances, and sound speed profiles. [Research supported by the U.S. Army Armament, Research Development and Engineering Center (ARDEC).]

5aPA15. Proper orthogonal decomposition for sensitivity analysis and classification of outdoor sound propagation. Chris Pettit (Aerosp. Eng. Dept., U.S. Naval Acad., 590 Holloway Rd., MS 11-B, Annapolis, MD 21402, pettitcl@usna.edu) and D. Keith Wilson (U.S. Army Cold Regions Res. and Eng. Lab., Hanover, NH 03755)

Simulations of outdoor sound propagation are compromised by many sources of uncertainty and error. A key step in validating computational acoustics models, which may be interpreted as the act of assessing predictive skill, is estimating ensemble sound pressure variability due to uncertainties in the parameters that define the propagation conditions. This paper describes elementary steps toward achieving this goal. The atmospheric surface layer is represented through Monin-Obukhov similarity theory and the acoustic ground properties with a relaxation model. Randomness is assumed across appropriate ranges of the governing parameters. The parameters are modeled herein as independent random variables, but future efforts will extend this formulation to include random field models of the parameters with varying levels of correlation. Sound propagation is predicted with the parabolic equation method. Latin hypercube sampling (LHS) and proper orthogonal decomposition (POD) are employed to develop low-dimension representations, i.e., reduced order models, of the sound pressure random field. Sensitivity of the sound pressure field is studied via the model-induced sensitivity of the POD mode coefficients to the system parameters as well as statistical dependence on the number of LH samples. Attention is also given to connecting the POD coefficients to documented indicators of refraction conditions.

5aPA16. Analysis of the effect of small currents on ray stability and caustic formation in a layered moving fluid medium. David Bergman (Lockheed Martin Maritime Systems and Sensors, 199 Borton Landing Rd., Moorestown, NJ 08570)

In studies of underwater sound it is usually sufficient to treat the medium of propagation as a static inhomogeneous medium. Including current and time-dependent changes adds to the overall understanding of the behavior of sound in underwater environments. The quiescent medium approach combines the movement of the fluid with the local sound speed to create an effective sound speed profile. Recently it has been demonstrated that the effects of small-scale fluid motion on the propagation of sound in the ocean can be of the same order of magnitude as that of the static sound speed profile [J. A. Colosi, J. Acoust. Soc. Am. **119**, 705–708 (2006)]. The stability of the acoustic ray equation in a moving fluid medium depends on the first and second derivatives of the sound speed profile and fluid velocity [D. R. Bergman, Waves Random Complex Media **15**(4), 417–435 (2005)]. In this presentation the effects of weak but rapidly changing currents on the stability and intensity of the acoustic field are investigated for environments with low Mach number. The main result will show that in real underwater environments small currents cannot be ignored when modeling or analyzing the propagation of sound.

5aPA17. Complex drag coefficient and complex Nusselt number for oscillatory flow. Jin Liu (United Technologies Res. Ctr., East Hartford, CT 06108) and Steven L. Garrett (Penn State Univ., State College, PA 16804)

Characterizing and choosing the optimal porous material for efficient heat transfer while minimizing viscous loss is important for improving the efficiency of thermoacoustic machines. The pressure drop for oscillating flow through porous media has both a viscous (real) part and an inertial (imaginary) part, making the drag coefficient for oscillatory flow a complex number [J. Liu, J. Acoust. Soc. Am. **115**, 2380 (2004)]. Similarly, a phase difference exists between the oscillating pressure and the oscillating temperature. A complex Nusselt number is therefore required to characterize heat transfer for oscillating pressure. For the first time, the complex Nusselt number will be related to the thermoviscous f-function favored by thermoacousticians. This identification also allows the heat transfer coefficient for oscillating flow to be determined experimentally, at sufficiently low amplitudes, without requiring the measurement of temperatures or the use of heat exchangers. At low frequencies, the real part of the complex Nusselt number converges to the steady-state (dc) heat transfer coefficient. At high frequencies, it recovers the results published by other researchers interested in the behavior of heat exchangers used in oscillating flows. [Work supported by the ONR.]

Session 5aPP

Psychological and Physiological Acoustics: Miscellaneous: Pipe Organs to T.V. Ads

Peter Zhang, Cochair

Michigan State Univ., Dept. of Physics and Astronomy, East Lansing, MI 48824

Minoru Tsuzaki, Cochair

Kyoto City Univ. of Arts, Faculty of Music, 13-6 Kutsukake-cho, Oe, Nishikyo-ku, Kyoto 610-1197, Japan

Chair's Introduction—8:30

Contributed Papers

8:35

5aPP1. The significance of relative loudness in listener identification of pipe organ stops. Alastair C. Disley and David M. Howard (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK)

As part of a wider study of the factors affecting listener identification of pipe organ stops, the authors' previous research [A. C. Disley and D. M. Howard, "Onset transient significance in listener identification of pipe organ stops," J. Acoust. Soc. Am. **119**, 3333, abstract only (2006)] has suggested that the relative loudness of a stop as presented to a listener may have a significant effect on its identification. Recordings of a representative selection of pipe organ stops were played to a group of 43 organists in a formal listening test. Each stop was presented twice, once at its original amplitude and once at an increased amplitude equal for all stops. Listeners were asked to classify each example using nine general categories of organ stops. The results suggest that increased amplitude improves listener recognition for some stops (particularly open flute stops), but produces significant misidentification of other stops, especially those with the greatest difference between the original and increased amplitude. These results suggest that caution should be employed when considering equalizing sample amplitudes in similar listening experiments.

9:05

5aPP3. Remembering what's missing: Modeling phoneme restoration and other complex forms of auditory induction as the triggering of a memory. Justin Aronoff (Neurosci. Program, Univ. of Southern California, HNB 27, Los Angeles, CA 90089, aronoff@usc.edu)

Although phoneme restoration is often spoken about and modeled as a unique phenomenon specific to language, the effect is very similar to restoration effects found with well-learned musical melodies. This suggests that both of these effects may be the result of a natural characteristic of learned temporal patterns, rather than a special mechanism. A recurrent connectionist network was designed to test this hypothesis. The network contained ten input and output nodes, which represented notes in a melody. The model was trained on sets of simple melodies, each consisting of eight randomly selected sequential notes. To test that the model was learning and identifying the melodies, the network was trained to predict the next note in the tune, which could only be accomplished by maintaining a representation of the sequence of prior notes. After the network learned the melodies, the fourth sequential note in each tune was removed and replaced by activation across all the input nodes, simulating noise. The model demonstrated accurate restoration, performing significantly better than chance, both at the noise position and over the notes that followed, suggesting that sequence learning may be sufficient to explain phoneme restoration and other forms of complex auditory induction.

8:50

5aPP2. Influence of attack transient and decay times of percussive sounds on the echo threshold. Hari V. Savitala and Jonas Braasch (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, hari.savitala@gmail.com)

The human ability to localize a direct sound source in the presence of reflected sounds is well known as the precedence effect. The echo threshold defines the time delay of the reflection, above which the reflection becomes audible as a separate event. For musical instruments, various attack transient and decay times affect the echo threshold, but often a constant echo threshold of 50 ms is assumed for music. The aim of this study is to determine in more detail how the parameters of a test sound affect the echo threshold. For this purpose, 12 adult listeners with normal hearing participated in a psychoacoustic experiment to determine the echo threshold for various percussive and nonpercussive instrument sounds. In the headphone-based experiment, the lead was placed with an interaural time difference (ITD) of ± 300 μs, while the lag was presented from the opposite side. Interstimulus interval (ISI) values of 0, 1, 2, 5, 10, 25, 50, and 100 ms were tested. The results of the experiment showed a large variation of the echo threshold. The smallest value of 7.50 ms was found for the castanet and the largest of 75.00 ms for a free-reed organ pipe.

9:20

5aPP4. Auditory priming releases Chinese speech from informational masking. Zhigang Yang, Jing Chen, Xihong Wu, Yanhong Wu (Dept. of Psych., Natl. Key Lab. on Machine Percept., Peking Univ., Beijing, China, 100871), Bruce Schneider, and Liang Li (Univ. of Toronto at Mississauga, Mississauga, ON, Canada L5L 1C6)

Before an English speech sentence is presented, hearing or reading the sentence without the last key word improves recognition of the last key word if the full-length speech sentence is presented under speech masking but not under noise masking. This phenomenon suggests a content priming effect on releasing speech from informational masking. To determine whether the priming effect extends to tonal Chinese speech, and, in particular, whether it can be induced by the target talkers voice, in the present study, listeners were presented with either same-voice/different-sentence primes or same-voice/same-sentence primes before hearing the target sentence in either two-talker-speech masking or noise masking. Under speech masking, each of the two prime types significantly improved recognition of the last key word in the full-length target sentence, but the content priming is stronger than the voice priming. Under noise masking, same-voice/same-sentence primes had a weak but significant priming effect, but same-voice/different-sentence primes had only a negligible priming effect. These results suggest that both content and voice cues can be used by listeners to release Chinese speech from informational masking, but only

3341   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3341**

5a SAT. AM

content cues are useful for releasing Chinese speech from energetic masking. [Work supported by China NSF and Canadian IHR.]

5aPP5. Auditory memory for frequency in sequential interference. Jeffrey J. DiGiovanni and Dennis T. Ries (School of Hearing Speech and Lang. Sci., Ohio Univ., Athens, OH 45701)

Percepts within working memory are weakened by the presence of intervening stimuli that lie along the same perceptual dimension. To this end, the effect of sequential interference on the storage of frequency information within auditory working memory was measured. The difference limen for frequency was measured by the method of constant stimuli for two diotic and two dichotic conditions. The first diotic condition served as a control and consisted of two 300-ms stimuli separated by 5 s. The frequency of the referent (first tone) was roved over a 100-Hz range centered at 435 Hz. Listeners indicated whether the target (second tone) was higher or lower in frequency than the referent. The second diotic condition added four interfering tones between the referent and the target. In the third condition, the interferers and referent/target pair were lateralized to opposite ears via the Stenger effect. In condition four, the phase of the interferers differed between ears by 180 deg. Results indicate (1) the addition of interferers increased the difference limen for frequency, and (2) when the interferer and target/referent pair differed along a single localization dimension, the effect of the interferers was reduced.

5aPP6. Speech recognition in coherently-amplitude modulated noise. Imran Dhamani, K. Naveen, and B. Rajashekhar (Dept. of Speech and Hearing, MCOAHS, MAHE, Manipal, Karnataka, India, imrandhamani@yahoo.co.in)

Comodulation masking release is a phenomenon that improves the detectability of a masked pure tone or speech signal by addition of a coherently amplitude-modulated energy above and/or the signal frequency. The majority of the studies done on comodulation masking release have studied threshold detection for a pure tone, or speech identification at or near threshold levels with a favorable signal-to-noise ratio. Moreover, various studies examining comodulation masking release under conditions of reduced frequency selectivity in listeners with cochlear hearing impairment indicate absence or reduced comodulation of masking release effect (Hall et al., 1984; Hall and Grose, 1985; Moore et al., 1993). In the current study, the speech recognition task was studied for the CMR effect at supra-threshold level and unfavorable signal-to-noise ratio of -10 dB. Results indicated absent CMR effect at supra-threshold level for the word recognition task for normal-hearing individuals.

5aPP7. Modulation detection thresholds for cochlear implants: Dependence on stimulation site and stimulus level. Bryan E. Pfingst, Rose A. Burkholder, Catherine S. Thompson (Kresge Hearing Res. Inst., Dept. of Otolaryngol., Univ. of Michigan, Ann Arbor, MI 48109-0506, bpfingst@umich.edu), and Li Xu (Ohio Univ., Athens, OH 45701)

Amplitude modulation of pulse trains is the strategy of choice for transmission of temporal-envelope information in most modern auditory prostheses. Previous studies by Fu, Colletti, and Shannon have shown high negative correlations between modulation detection thresholds (MDTs) (measured at only one site in the electrode array) and speech recognition with cochlear or brainstem auditory prostheses. The objective of this study was to characterize MDTs across the parameter space used in cochlear prosthesis speech processors: all usable stimulation sites at multiple stimulus levels. Across-site patterns of MDTs varied considerably from subject to subject. In general, sites with the poorest MDTs were located in the basal third of the scala-tympani electrode array, but this pattern did not hold for all subjects or all stimulus conditions. The mean MDT for two levels (30% and 70% of dynamic range) was highly correlated with the

mean for five levels. Mean MDTs at these two levels averaged over four adjacent electrodes in any of five segments of the electrode array were negatively correlated with speech recognition scores. This study suggests that in order to adequately characterize a subject's modulation detection acuity, it is necessary to sample MDTs at multiple sites and levels. [Work supported by NIH/NIDCD.]

5aPP8. Lateralization of sine tones—Interaural time versus interaural phase. Peter X. Zhang and William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., 4230 BPS Bldg., East Lansing, MI 48824)

Five listeners estimated the lateral positions of 50 sine tones in a headphone experiment designed to determine whether the human sense of sound location correlates better with the interaural phase difference (IPD) or the interaural time difference (ITD). In any experimental block the IPD values ranged from -150 to $+150$ degrees, and the frequencies were chosen such that the ITDs ranged from -1000 to $+1000$ microseconds. The frequencies were all in the range where human listeners are known to be able to lateralize tones based on the ITD in the waveform fine structure. It was found that the lateralization responses correlated with the ITD much better than with the IPD. The average variance was five times smaller for the ITD hypothesis compared to the IPD hypothesis, and only the ITD hypothesis led to a well-fitting (compressive) function. For the ITD function, individual compressive exponents varied considerably and averaged 0.75. For the IPD function the exponents were too small to be meaningful. Comparison with previous lateralization studies indicates the importance of presenting the entire range of stimulus parameters in all experimental blocks because it appears that listeners tend to use the entire range of allowable responses in any block. [Work supported by the NIDCD.]

5aPP9. Influence of visual stimuli in the sound quality evaluation of loudspeaker systems. Alex Karandreas and Flemming Christensen (Sound Quality Res. Unit, Dept. of Acoust., Aalborg Univ., Fredrik Bajers Vej 7 B5, DK-9220 Aalborg, Denmark)

There is currently an increasing demand to evaluate sound quality attributes of products and to understand to what extent they influence a user's overall impression, since there is usually more than one modality stimulating this evaluation. The present study uses loudspeakers as an example and evaluates the overall impression in relation to hearing and vision. In order to quantify the bias that the image of a loudspeaker has on the sound quality evaluation done by a naive listening panel, loudspeaker sounds of varied degradation are coupled with positively or negatively biasing visual input of real loudspeakers, and in a separate experiment by pictures of the same loudspeakers. In order to choose loudspeakers that provide a sufficient range of visual bias a preliminary visual-only experiment has been conducted. From the ongoing experiments it is possible to evaluate how much the auditory perception of a loudspeaker can be biased from visual input and study how the two modalities interact. Results from the experiments are presented.

5aPP10. Head-related transfer functions via the fast multipole accelerated boundary element method. Nail A. Gumerov, Ramani Duraiswami, and Dmitry N. Zotkin (Perceptual Interfaces and Reality Lab., Inst. for Adv. Comput. Studies, Univ. of Maryland, College Park, MD 20742)

Computation of head-related transfer functions (HRTF) via geometrical meshes and numerical methods has been suggested by a number of authors. An issue facing this approach is the large computational time needed for high frequencies, where the discretization must include hundreds of thousands of elements. Conventional computational methods are unable to achieve such computations without excessive time or memory. We use a newly developed fast multipole accelerated boundary element

method (FMM/BEM) that scales linearly both in time and memory. The method is applied to the mesh of the widely used KEMAR manikin and its HRTF computed up to 20 kHz. The results are compared with available experimental measurements.

11:20

5aPP11. Perceptual measurement of sound level differences between TV programs and advertisements. Eiichi Miyasaka (Musashi Inst. of Technol., 3-3-1, Ushikubo-Nishi, Tsuzuki-ku, Yokohama, Japan) and Takahiro Kamada (Pioneer Co., Tsurugashima, Saitama, Japan)

In order to investigate loudness difference between TV program materials and advertisements (CMs) inserted in the materials, sound levels of CMs broadcast in five terrestrial broadcasting stations throughout a day were quantitatively measured and a perceptual experiment was performed. The averaged sound levels (ASLs) and the standard deviations (SDs) of 4262 CMs were concentrated on -7 dB and 5 dB, respectively. Next, three types of CMs with different ASLs and SDs were used to the perceptual experiment. These ASLs and SDs were (-5.1 and 1.8 dB for CM-1), (-7.1 and 4.9 dB for CM-2), (-14.2 and 10.1 dB for CM-3), respectively. The reference sound as a main program material was speech uttered by a NHK female announcer with duration of 10 s. The ASL and the SD were -6.5 and 10.1 dB, respectively. The transformed up-down method was used for 12 observers with normal hearing. The results show that all CMs were perceived at least 3 to 5 dB louder than the reference for a half of the observers. These results show that there is large perceptual discrepancy between CMs and the main program materials. [Work supported by The Okawa Foundation.]

SATURDAY MORNING, 2 DECEMBER 2006   HONOLULU ROOM, 8:30 TO 11:30 A.M.

Session 5aSA

Structural Acoustics and Vibration: Numerical Methods in Structural Acoustics and Vibration

Dean E. Capone, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Toshimitsu Tanaka, Cochair
Kobe Steel, Ltd., Mechanical Engineering Research Lab., 1-5-5 Takatsukadai, Nishi-Ku, Kobe City, Hyogo 651-2271, Japan

Invited Papers

8:30

5aSA1. Recent developments and applications of energy finite- element analysis. Kuangcheng Wu (Dept. of Signatures and Hydrodynamics, Northrop Grumman Newport News, Newport News, VA 23607) and Nickolas Vlahopoulos (Univ. of Michigan, Ann Arbor, MI 48109)

The energy finite-element analysis (EFEA) comprises a finite-element-based solution for vibration and acoustic modeling that uses energy-based variables for formulating the governing differential equations. This presentation covers the main points in the derivation of energy finite-element analysis (EFEA), validation through comparison with test data, and case studies that demonstrate the computational capabilities of the EFEA. The EFEA can be applied in various engineering disciplines as it will be demonstrated by the case studies. Thus, it comprises a general purpose simulation method for vibration and acoustic analysis of complex systems. [Work support for this research is provided by ONR code 331.]

8:50

5aSA2. Fast finite-element analysis for damping of automotive structures having elastic bodies, viscoelastic bodies, porous media, and gas. Takao Yamaguchi (Dept. of Mech. System Eng., Gunma Univ., 1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515, Japan) and Yoshio Kurosawa (Fuji Heavy Industries, ltd., Sano-shi, Gunma, 327-0512 Japan)

A numerical method is proposed to calculate damping properties for automotive soundproof structures involving solid bodies, porous media, and air in two-dimensional regions. Both effective density and bulk modulus have a complex quantity to represent damped sound fields in the porous media. Particle displacements in the media are discretized using a finite-element method. For damped solid bodies, displacements are formulated using conventional finite elements including complex modulus of elasticity. Displacement vectors as common unknown variables are solved under coupled condition between solid bodies, porous media, and gas. Further, by applying an asymptotic method to a complex eigenvalue problem, explicit expressions of modal loss factor for the mixed structures are derived. The proposed methods yield appropriate results for some typical problems and this method diminishes computational time for large-scaled finite-element models. Moreover, it is found that damping can be coupled in the mixed structures. An expression to calculate a share of dissipated energy for each element in mixed structures is also derived. Damping behaviors in sound bridge phenomena are analyzed using the proposed method.

5a SAT. AM

5aSA3. Vibro-acoustic analysis using hybrid finite element, boundary element, and statistical energy analysis models. Phil Shorter and Vincent Cotoni (ESI US R&D, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130, pjs@esi-group-na.com)

This paper presents an overview of the theory and application of the hybrid FE-SEA method. The method provides a rigorous way to model the fully coupled response of a vibro-acoustic system across a broad frequency range using a combination of FE, BEM, and SEA. The theory of the method will be discussed and results from the following applications will be presented: (i) modeling the transmission loss and radiation efficiency of trimmed automotive subassemblies; (ii) creating efficient models of structure-borne noise in an automotive body-in-white; (iii) modeling structure-borne noise transmission in a commercial aircraft fuselage; and (iv) modeling the response of a satellite antenna to broadband acoustic loading.

Contributed Papers

9:30

5aSA4. Simulation of borehole acoustic measurements with adaptive finite elements for the accurate and efficient assessment of rock formation properties. Christian Michler, Leszek Demkowicz, and Carlos Torres-Verdin (Inst. for Comput. Eng. and Sci., Univ. of Texas, 201 East 24th St., Austin, TX 78712, c.michler@ices.utexas.edu)

Borehole acoustic measurements are routinely used to probe *in situ* properties of rock formations. The numerical simulation of such measurements involves numerous outstanding challenges, such as the modeling of large spatial variations of rock porosity and anisotropy, material diffractions, source-sensor effects, and the need for a reflectionless truncation of the computational domain. To simulate borehole acoustic measurements, a new axisymmetric finite-element formulation in the frequency domain has been developed and successfully tested. The formulation couples acoustic phenomena within the fluid-filled borehole with elasticity in both the probed rock formations and the measurement tool. One of the central components of our formulation is the automatic adaptivity of element size and approximation order of the finite-element spatial discretization. No user interaction is necessary to control the adaptivity. Such a formulation enables accurate simulation results in the presence of large contrasts of material properties. A novel combination of the perfectly matched layer enhanced with finite-element adaptivity effectively truncates the computational domain. Several challenging benchmark problems are presented that confirm the accuracy, reliability, and efficiency of our method to reproduce complex waveforms excited by monopole and dipole sources in a fluid-filled borehole that penetrates layered media.

9:45

5aSA5. Perfectly matched layers for the numerical approximation of the radiation boundary condition in structural acoustic finite element tools. Mario Zampolli, Alessandra Tesei, Finn B. Jensen (NATO Undersea Res. Ctr., La Spezia, Italy), and John B. BlottmanIII (Naval Undersea Warfare Ctr., Newport, RI 02841)

The Berenger perfectly matched layer (PML) is one of a variety of numerical techniques for approximating the Sommerfeld radiation boundary condition. In acoustics, the PML is a nonphysical finite thickness layer of fluid or solid material that surrounds the physical computational domain of interest and acts as an anisotropic absorber of the outgoing waves. The absorption coefficients in the PML are such that the acoustic energy is dissipated selectively only in the direction perpendicular to the interface between the PML and the physical domain. The study presented here shows that the appropriate choice of PML absorbing functions can have a beneficial effect on the convergence of frequency domain acoustic finite-element tools. In the context of elastic target scattering computations, it is shown how the PML can be applied in direct contact with the wet surface of convex elastic targets. Other examples from low-frequency geoacoustic benchmarking applications demonstrate the adaptability of the PML to a variety of geometries and to problems where the radiation condition must be imposed across inhomogenous layers. The numerical results obtained with the finite-element/PML tool are compared to results from finite-element/infinite-element codes and to results from parabolic equation models.

10:00–10:15   Break

10:15

5aSA6. Reconstruction of vibro-acoustic response of a plate using Helmholtz equation least-squares method. Huancai Lu and Sean F. Wu (Mech. Eng. Dept., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202, huancai.lu@gmail.com)

A numerical investigation of reconstructing the vibro-acoustic responses of an arbitrary structure subject to vibration excitations based on the Helmholtz equation least-squares (HELS) method is presented. It is emphasized that in many engineering applications, the exact solution to a general vibrating structure does not exist, and the HELS method is one way of getting approximate solutions in a cost-effective manner. In this study, the test object is a simply supported, unbaffled thin plate. The reason for selecting this simply supported plate is that the analytic solutions to the plate vibrations are readily available. The field acoustic pressures generated by the Rayleigh integral are taken as input to HELS algorithms to reconstruct the normal velocity and normal acoustic intensity on the plate surface using Tikhonov regularization associated with generalized cross-validation methods. The reconstructed normal surface velocities are compared with the benchmark values, and the out-of-plane vibration patterns at the first five natural frequencies are compared with natural modes of the simply supported plate. The effects of measurement aperture, stand-off distance, and location of the origin of the coordinate system on the resultant accuracy of reconstruction are examined. [Work supported by NSF.]

10:30

5aSA7. Application of the short-time Fourier transform and the two-dimensional Fourier transforms to detect damage in plates. Jamal Assaad, Faysal El Youbi, Sebastien Grondel, and Emmanuel Moulin (IEMN-DOAE, 59313 Valenciennes, cedex 9, France)

Lamb waves are widely used in the fields of characterization and non-destructive testing of structures. Usually, the full characterization of the structures requires two kinds of information: the first one is related to the material quality, the other one aims at the evaluation of the structure health, i.e., the detection of localized damage or defects. Time-frequency analyses or two-dimensional Fourier transforms (2DFT) have been used separately to solve such a problem. The purpose in this work is therefore to present a dual signal processing approach based on the relationship between the STFT (short-time Fourier transform) and the 2DFT amplitudes that allows a better interpretation and identification of Lamb modes. This relationship will be numerically verified by a two-dimensional finite-element method. Experimentally, a 3-mm-thick aluminum plate with an emitter and a receiver are considered. The emitter consists of two piezo-electric transducers while the receiver consists of a 32 multi-electrode piezoelectric transducer. This receiver has been developed especially to increase signal-to-noise ratio at the reception. In order to simulate damages a hole with different diameter was introduced in the plate center. The application of this dual signal processing has located conversion modes phenomenon.

10:45

5aSA8. Sound radiation from finite plates excited by a space-time stream of random impulses. Katsuji Akamatsu (Machinery Acoust., 1-1-2-314 Obanoyama Shinohara, Nada-ku, Kobe 657-0015, Japan, akamatsu@s4.dion.ne.jp), Takao Yamaguchi (Gunma Univ., Kiryu 376-8515, Japan), and Junichi Kanazawa (Musashino Damping Technol., Tachikawa 190-0002, Japan)

Sound radiation from baffled finite elastic plates that are subjected to impulses occurring at random points on the surface, at random intervals, and with random strength is analyzed. The plate response is obtained in the Stieltjes integral form of the plate impulse response function with respect to the occurrence of impulses. The total sound power radiated from the plate is formulated according to Heckl's approach. Explicit expressions for the expected value of the plate response and the radiated sound power are derived for the case in which the stream of impulses is uncorrelated. An approximate solution for the radiated sound power is obtained by assuming light damping of plates and by neglecting modal coupling effects. For comparison, the exact and approximate solutions are evaluated numerically for a plate with constant loss factors. The analysis is applied to the prediction of rainfall noise by expressing the expected value of the exciting force in terms of the size of raindrops, their terminal velocity, and the rainfall rate.

11:00

5aSA9. Dynamic analysis of shuttle bus generated vibration in an airport rental car center with an elevated roadway. James Phillips (Wilson, Ihrig & Assoc., 5776 Broadway, Oakland, CA 94618, jphillips@wiai.com)

A finite-element analysis (FEA) model was developed for a new rental car center at an international airport in the United States. Incorporated into the structure was an elevated roadway for shuttle buses transporting rental car customers. Of concern was possible vibration transmitted into occupied portions of the structure from shuttle buses. Measurements were conducted on an existing elevated roadway at the airport in order to determine dynamic force inputs to be applied in the FEA model developed to predict the vibration response inside the rental car center. Impulse response measurements were conducted and then the vibration response of the roadway was measured during shuttle bus passbys. The results of these tests were used to derive the dynamic forces generated by the moving buses on the structure. A finite-element analysis (FEA) model of the elevated roadway was also developed. This model was used to verify the derived dynamic force levels and to verify the modeling approach applied to the rental car center.

11:15

5aSA10. Parametric relations for porous foams and their use in optimization of structural acoustic performance of multilayer trim components. Eleonora Lind and Peter Goransson (KTH Aeronautical and Vehicle Eng., MWL, SE-10044 Stockholm, Sweden, eleonora@kth.se)

Lightweight porous acoustic multilayer trim components have traditionally been specified in terms of sound absorption and sound transmission loss performance targets. Importantly, the material specification of the trim component developed only for absorption and sound transmission loss may be suboptimal in terms of, e.g., sound radiation behavior. This highlights the necessity for accurate, computationally efficient, and robust simulation method, which should form an integral part of a multidisciplinary optimization tool. In addition, for such optimization to be physically meaningful the design parameters used should be based on relations between microstructural dimensions and properties, and the corresponding macroscopic parameters describing porous materials. This paper discusses recent results from research focused on the parametrization of porous foams, i.e., continuous links between micro-dimensions, elasticity, density, and flow resistivity, used in higher order 3-D finite element simulations of multilayer components for structural-acoustic applications. In addition to the microdimensions, also the thicknesses of individual layers have been used as design parameters, with the overall weight as the design objective. Acoustic and vibration targets, velocity of the radiating surface, radiation efficiency, transmission, etc., as well as upper and lower bounds of the microdimensions, have been used as constraints.

Session 5aSCa

Speech Communication: Speech and Hearing in Targeted Populations (Poster Session)

Mini Shrivastav, Cochair

Univ. of Florida, Dept. of Communications Science and Disorders, 336 Dauer Hall, Gainesville, FL 32611

Hideki Kasuya, Cochair

International Univ. of Health and Welfare, Dept. of Speech, Language and Hearing Science, 2600-1 Kita-kanemaru, Otawara 324-8501, Japan

Contributed Papers

All posters will be on display from 8:00 a.m. to 11:00 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 9:30 a.m. and contributors of even-numbered papers will be at their posters from 9:30 a.m. to 11:00 a.m.

5aSCa1. Making young ears old and old ears even older: Simulating a loss of synchrony. Ewen MacDonald, Kathy Pichora-Fuller, Bruce Schneider, and Willy Wong (Univ. of Toronto, Toronto, ON Canada M5S 3G9)

Age-related changes in the auditory system have been attributed to three independent factors: OHC damage, changes in endocochlear potentials, and loss of neural synchrony. In previous studies, a jitter algorithm has been used to simulate the loss of synchrony in young adults (Mac-Donald *et al.*, 2005). In this study, the effect of jitter on old adults with good audiograms in the speech range is explored. SPIN-R sentences were presented in two SNR and three processing conditions: intact, jitter, and smear. The parameters of the jittering algorithm were the same as those used with young adults. The parameters of smearing algorithm were chosen to match the spectral distortion produced by jitter algorithm. While both the jitter and smear conditions resulted in a significant decline in word identification, the decline was largest in the jitter condition. Psychometric functions were fitted to the data and compared to previous work with young adults. The comparison supports the hypothesis that loss of synchrony can adversely affect speech intelligibility in noise, and is consistent with the hypothesis that loss of synchrony occurs with age. As well, the comparison suggests that the effect of jitter may be linear.

5aSCa2. Adult age differences in the use of envelope cues to identify noise-vocoded words with a varying number of frequency bands. Signy Sheldon, Kathy Pichora-Fuller, and Bruce Schneider (Univ. of Toronto at Mississauga, Rm. CCIT 4163, 3559 Mississauga Rd., Mississauga, ON, L5L1C6, Canada, signy.sheldon@utoronto.ca)

Older adults with good audiograms have difficulty understanding speech in noise. Age-related differences have been found on some temporal processing measures such as gap detection; however, older adults are believed to have well-preserved ability to use envelope cues to identify words. Following Shannon *et al.* (1995), we used noise-vocoded speech such that the amplitude envelope of speech was retained in frequency bands but filled with noise, thereby obliterating fine structure cues within each band. In experiment 1, younger and older listeners heard a list of words. Each word was presented first with one vocoded frequency band, and the number of bands was incremented until the listener correctly identified the word. The average number of bands required for correct identification was found to be identical for both age groups. In experiment 2, both age groups identified words in four blocked noise-vocoded conditions (16, 8, 4, and 2 bands). Younger adults outperformed older adults. Al-

though older adults were as able as younger adults to use envelope cues cumulatively in experiment 1, they were less able to use these cues without the benefit of repetition.

5aSCa3. Audiovisual perception of voicing with age in quiet and cafe noise. Dawn Behne (Dept. of Psych., Norwegian Univ. of Sci. and Technol., 7491 Trondheim, Norway), Yue Wang (Simon Fraser Univ., Burnaby, BC V5A 1S6, Canada), Magnus Alm, Ingrid Arntsen, Ragnhild Eg, and Ane Valsø (Norwegian Univ. of Sci. and Technol., 7491 Trondheim, Norway)

Research has shown that voicing is difficult to discern in noisy environments. While voicing may be difficult to resolve from visual cues, acoustic cues for voicing are relatively robust. This study addresses these factors with normally aging audiovisual perception. Identification responses were gathered with 19–30-year-old and 49–60-year-old adults for audiovisual (AV) CVs differing in voicing and consonant place of articulation. Materials were presented in quiet and in cafe noise (SNR=0 dB) as audio-only (A), visual-only (V), congruent AV, and incongruent AV. Results show a tendency toward use of visual information with age and noise for consonant place of articulation. Notably for voicing, incongruent AV materials that had one voiced component, regardless if it was A or V that was voiced, were consistently perceived as voiced in both age groups and regardless of noise. Only if the A and V components were both voiceless was the syllable perceived as voiceless. These findings indicate the influence of age and noise in the use of perceptual information to identify place of articulation. That voicing is robustly salient from either audio or visual information, despite the unlikely presence of strong visual cues for voicing, indicates a possible bias toward the perception of voicing.

5aSCa4. Energy suppression of steady-state portions of vowels while maintaining the energy of consonants better improves speech intelligibility for elderly listeners in reverberation. Yusuke Miyauchi and Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554 Japan, m-yuusuk@sophia.ac.jp)

In a reverberant environment, overlap-masking renders speech perception difficult. Arai itet al. proposed that energy suppression of steady-state portions of speech improves speech intelligibility with young subjects [Acoust. Sci. Technol. 23, 229–232 (2002)]. Audibility degrades with age. Therefore, reverberation can be a more critical barrier to speech perception by elderly listeners. To investigate the effect of suppressing reverberation for elderly people, a listening test was conducted with 25 elderly

3346   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3346**

subjects (mean age: 73.3 years) using the following three types of speech: (a) speech without energy suppression of steady-state portion, (b) speech with the suppression of steady-state portions of speech, and (c) speech with the suppression only of vowels, while maintaining the energy of consonants. As expected, speech intelligibility by method (b) (46.7%) was markedly improved from (a) (42.2%), but method (c) (52.0%) improved it more. Because the energy of consonants is less than that of vowels, overlap-masking attendant on previous vowels would largely affect perception of subsequent consonants. These results suggest that suppressing the steady-state portions of vowels while maintaining the energy of consonants serves elderly persons well to improve speech intelligibility in a reverberant environment. [Work supported by JSPS.KAKENHI (16203041).]

5aSCa5. Older adult's identification and memory of synthetic and natural speech in noise. Candice Q. McCarty and Aimee Surprenant (Purdue Univ., 703 Third St., West Lafayette, IN 47907)

This research tested older adult's performance on identification and recall when presented sentences taken from the Speech Perception in Noise Test [Bilger *et al.*, J. Speech Hear. Res. **27**, 32–48 (1984)] synthesized using AT&T Natural Voices speech synthesizer. Performance on synthesized speech was compared to identification and recall of natural speech in noise. In experiment 1, participants were instructed to identify the final word of high- and low- predictability natural and synthesized sentences presented at three signal-to-noise ratios. The results showed that natural speech was easier to identify than synthesized speech at every speech-to-noise ratio for high and low predictability sentences. In addition, replicating Pichora-Fuller *et al.* [J. Acoust. Soc. Am. **97**, 593–608 (1995)], pure tone thresholds were significantly correlated with identification performance for both natural and synthetic speech. In experiment 2, participants were given a sentence span task in which they were presented two to five natural and synthesized sentences and asked to recall the final words. These experiments show that, even with high context, synthetic speech is more difficult to understand and remember than natural speech. These results highlight concerns about the use of synthesized speech in assistive devices, particularly for elderly listeners. [Work supported by NIH.]

5aSCa6. Spectral shape discrimination and speech identification in noise in elderly hearing-impaired listeners. Mini Shrivastav (Dept. of Commun. Sci. and Disord., Univ. of Florida, 336 Dauer Hall, Gainesville, FL 32611)

Elderly hearing-impaired listeners have difficulties in the recognition and discrimination of consonants, particularly those that share the same manner of articulation. An important cue that distinguishes the manner of articulation of these consonants is spectral shape. A previous study (Shrivastav *et al.*, 2006) found a moderate predictive relation between spectral-shape discrimination thresholds and syllable identification scores of elderly hearing-impaired listeners, when all stimuli were presented in quiet. The present study examined the contribution of spectral-shape discrimination abilities to speech-identification performance of elderly hearing-impaired listeners in the presence of background noise. The study included a group of elderly hearing-impaired listeners, with a group of young normal-hearing adults included for comparison purposes. Listeners were tested on a series of speech-identification and spectral-shape discrimination tasks, while ensuring that all stimuli were at least minimally audible to all the hearing-impaired listeners. The contribution of spectral-shape discrimination abilities to speech-identification performance in background noise was examined and compared to the results in quiet for young normal-hearing and older hearing-impaired listeners.

5aSCa7. The use of subsegmental information in sentence comprehension with or without formant transitions by normal-hearing and hearing-impaired listeners. Jae Hee Lee and Diane Kewley-Port (Dept. of Speech and Hearing Sci., 200 S. Jordan Ave., Bloomington, IN, 47405, jaejalee@indiana.edu)

Using a noise replacement paradigm, Lee and Kewley-Port [J. Acoust. Soc. Am., **119**, (2006)] examined sentence comprehension by young normal-hearing (YNH) and elderly hearing-impaired (EHI) listeners when sentences were processed to present only subsegmental information (i.e., either steady-state or formant transitions) in sentences processed in four different ways. Results showed that correct word responses by EHI listeners were more affected by the type of information in the four conditions compared to YNH listeners. To compare the use of subsegmental information between listener groups who were carefully matched for audibility and age, new analyses of correctly identified phonemes and different word-error patterns were made for both correct and incorrect word responses. Correlation analyses examined the relations among correct and incorrect word responses, phoneme scoring, hearing thresholds, and age. Despite a larger number of incorrect answers by EHI listeners, word-error patterns, as well as an ability to extract phonemes using subsegmental information, were similar between YNH and EHI listeners. Analyses for EHI listeners showed correct word responses were strongly correlated with correct phoneme identification and with hearing thresholds, but not with age. [Work supported by NIHDCD-02229.]

5aSCa8. Auditory-visual integration and lipreading abilities of older adults with normal and impaired hearing. Mitchell S. Sommers (Dept. of Psych., Washington Univ., Campus Box 1125, St. Louis, MO 63130, msommers@wustl.edu), Nancy Tye-Murray, and Brent Spehar (Washington Univ. School of Medicine, St. Louis, MO 63130)

The current study investigated how age-related hearing impairment affects lipreading and auditory-visual integration in older (above age 65) individuals. The performance of 53 normal-hearing and 24 mild-to-moderate hearing-impaired older adults was compared on auditory-only (A), visual-only (V), and auditory-visual (AV) speech perception, using consonants, words, and sentences as stimuli. All testing was conducted in the presence of individually specified multi-talker background babble, to obtain approximately equivalent A-only performance across the groups. In addition, we compared normal-hearing and hearing-impaired individuals on measures of auditory enhancement, visual enhancement, and auditory visual integration that were derived from the A, V, and AV performance scores. In general, normal-hearing and hearing-impaired older adults performed similarly on measures of visual-only and auditory-visual speech perception. The one exception to this finding was that hearing-impaired adults performed significantly better than normal-hearing participants on visual-only identification of words. Measures of visual enhancement, auditory enhancement, and auditory-visual integration did not differ as a function of hearing status. Overall, the results of the current study suggest that despite increased reliance on visual speech information, hearing-impaired older adults do not exhibit better visual-only speech perception or auditory-visual integration than age-matched normal-hearing individuals.

5aSCa9. Effects of intermodal timing difference and speed difference on auditory-visual speech perception. Akihiro Tanaka (Dept. of Psych., Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 113-0033 Tokyo, Japan, tanaka@L.u-tokyo.ac.jp), Shuichi Sakamoto, Komi Tsumura, and Yo-iti Suzuki (Tohoku Univ., Sendai 980-8577, Japan)

Previous studies have shown that lipreaders possess insufficient sensitivity to intermodal timing differences and adapt to the timing difference when it is kept constant. This study investigated effects of intermodal timing differences and speed differences on auditory-visual speech perception. We used 20 minimal pairs of Japanese four-mora words such as *mizuage* (catch landing) versus *mizuame* (starch syrup). We administered intelligibility tests to younger and older adults. Words were presented

under visual-only, auditory-only, and auditory-visual (AV) conditions. Effects of AV asynchrony by time-expanded speech (speed difference) on AV benefit (i.e., the superiority of AV performance over auditory-only performance) were compared with those by intermodal timing shift (timing difference). Results showed that the AV benefit declined as the timing difference and speed difference increased. Results also revealed that younger adults adapted to the timing difference but did not adapt to the speed difference. Older adults adapted neither to the timing difference nor to the speed difference. These results suggest that intermodal lag adaptation in younger adults requires a constant timing difference between a talker's moving image and voice. These results might be useful for design of a multimodal speech-rate conversion system.

5aSCa10. Description of mouth shapes used in traditional Japanese sign language by comparison to those of speech utterance. Yumiko Fukuda (Natl. Rehabilitation Ctr. for the Disabled, Tokorozawa, Japan, fukuda@rehab.go.jp) and Shizuo Hiki (Waseda Univ., Tokorozawa, Japan)

Mouth shapes used in traditional Japanese sign language were described systematically through comparison to those in uttered speech. A framework of mouth shape descriptions that had been proposed by the present authors included (1) projection of mouth shapes of Japanese vowels and semi-vowels to the lip articulation plane of the three-dimensional IPA vowel chart and (2) grouping of discriminating consonants in lip-reading on the IPA articulation matrix. It was transferred to a coordinate consisting of contractions of muscles involved in mouth movement. In traditional Japanese sign language, the mouth shape, when added as a part of facial expression, modifies meaning of the word conveyed by a hand gesture. Those mouth shapes were extracted from conversations among native signers and were symbolized by their respective degrees of closing, rounding, and pulling corners, and by categories borrowed from Japanese vowels and some consonants. Using the extended framework, those mouth shapes that were elected for visual discrimination ranged more widely than those elected to discriminate phonemes aurally. Video pictures were retrieved according to their symbols from an electronic dictionary of 250 basic sign words and 1500 exemplary sentences of their meanings, as edited by the Japanese sign language research group, NRCD.

5aSCa11. Audiovisual speech integration in children with autism spectrum disorders. Julia R. Irwin (Haskins Labs., 300 George St., New Haven, CT 06511)

Children with autism spectrum disorder (ASD) appear to be less influenced by visual speech information than typically developing children, as measured by their responses to mismatching auditory and visual (McGurk) stimuli. This study examined whether this reduction in sensitivity to the McGurk effect is due to eye gaze aversion, a hallmark of ASD. Children with ASD and typically developing controls (TD) were presented with videotaped consonant-vowel (CV) stimuli. Stimuli were digitally edited to create either an audiovisual (AV) match (AV /ma/ or /na/) or mismatch (audio /ma/ and visual /ga/). Responses were considered visually influenced if participants reported hearing /na/ for the mismatched stimuli. Using eye-tracking methodology, only trials where participant's gaze was fixated on the speaker's face during consonantal closure were included in analyses. Initial analyses reveal, when fixated on the speaker's face, children with ASD show significantly less visual influence relative to typically developing controls. It is proposed that the reduced influence of visual speech information in ASD is best understood at the level of visual speech processing, perhaps arising from a lack of experience with speaking faces. Implications for understanding visual speech processing in typical development and developmental disability are discussed. [Work supported by NIH grant DC-00403 (Cathi Best, PI.]

5aSCa12. Auditory temporal order judgments in children with specific language impairment and autism spectrum disorder. Sheryl Rosen, Mark De Ruiter, and Li Hsieh (Wayne State Univ., 207 Rackham, 60 Farnsworth, Detroit, MI 48202)

The Tallal repetition test (TRT) measures the ability of listeners to detect, associate, sequence, and remember complex patterns of stimuli (serial memory) The TRT contains sequencing and serial memory subtests with varying short and long interstimulus intervals (ISIs) Three groups of 12 children were tested using a modified version of the TRT. Subject groups consisted of 7- to 10-year old children with typical development (TD), autism spectrum disorder (ASD), and specific language impairment (SLI). The TRT contained both tone and CV syllable stimuli to assess temporal processing. As ISI increased, performance improved for both children with TD and SLI in both tone and syllable sequencing tasks, but not for children with ASD, especially in the tone task. Children with TD and those with ASD performed better than those with SLI across most subtests, especially in tasks with short ISIs. Analysis of performance in the serial memory portion of the TRT revealed a predicted pattern of decreased performance as stimuli elements increased. Overall, children with TD and those with ASD appear to have superior auditory temporal order judgment abilities when compared to children with SLI. The linguistic relevance factor shall be considered in assessment and training of various disordered populations.

5aSCa13. Perception of synthesized syllables and sentences by listeners with hearing loss. Corine Bickley, Amanda Dove, Andrea Liacouras, and Jolene Mancini (Dept. of Hearing Speech and Lang. Sci., Gallaudet Univ., 800 Florida Ave NE, Washington DC 20002)

The goal of this project is to create a set of synthesized speech materials that yields listening results equivalent to those that would be obtained using recorded naturally spoken materials, such as the CUNY Nonsense Syllable Test and the CASPERsent set of topic-specific sentences. One research interest of the RERC-HE is the following: Can synthesized speech be used as a substitute for recorded human speech in aural rehabilitation and in hearing aid research? The listeners in this study are both adults with normal hearing and with hearing loss. We have tested two listeners with normal hearing and four with hearing loss using nonsense syllables synthesized with DECtalk. The patterns of errors in their identification of phonemes follow expected patterns when subjects' responses were analyzed with respect to their audiometric profiles and the acoustic characteristics of the synthesized phonemes. The same group of listeners is now being tested using sentences from the CASPERsent lists synthesized with two commercially available text-to-speech systems: DECtalk, which uses a formant-type synthesizer, and ATT Natural Voices, which uses a concatenative synthesizer. Results of these listening tests will be presented. [This project is funded through the Rehab Engineering Research Center—Hearing Enhancement through the NIDRD.]

5aSCa14. Effects of dichotic listening for hearing impaired people. Ai Sasaki, Atsunobu Murase, Shuichi Sakamoto, Yôiti Suzuki (Res. Inst. of Electron. Commun., Grad. School of Info. Sci., Tohoku Univ. Katahira 2-1-1, Aoba-ku, Sendai, 980-8577 Japan, asasaki@ais.riec.tohoku.ac.jp), Tetsuaki Kawase, and Toshimitsu Kobayashi (Tohoku Univ., Aoba-ku, Sendai, 980-8577 Japan)

Methods applying the dichotic listening algorithm have been proposed and examined for application in hearing aids. Murase *et al.* (2005) proposed an algorithm in which dichotic listening was simply implemented by a set of complementary high-pass and low-pass filters. One filter is for either the right or left channel; the other filter is used for the other channel. This study examined two unsolved problems with this algorithm. One is the reason why this algorithm is effective. The other is probable degradation of sound localization that occurs because interaural level differences and interaural time differences is only slightly available with this signal processing. This problem is considerable when the sounds come from frontal incidence. Regarding the first problem, intelligibility tests were

performed using the low-frequency-boosted sounds. The intelligibility scores under the dichotic listening conditions should be higher than those under the diotic listening conditions if the effectiveness of this algorithm is attributable to the reduction of upward spread of masking. Results suggest that this dichotic listening reduced the upward spread of masking. Regarding degradation of performance of sound localization, sound localization test results suggest that time-intensity trading might recover degradation with dichotic listening algorithm.

5aSCa15. Effects of auditory training on hearing aid acclimatization. Jack M. Scott III (Texas Medical & Surgical Assoc., 8440 Walnut Hill Ln., Ste. 500, Dallas, TX 75231, jack.scott@texasmedsurg.com)

New hearing-aid users may experience a gradual improvement in speech performance over time that typically plateaus 6–12 weeks following hearing aid fitting. The delayed improvement may be due to neural reorganization of the auditory system following the introduction of amplified auditory input. Unfortunately, 30 days is the typical duration for most hearing aid trial periods; therefore, many individuals may be returning their hearing aids before being able to evaluate the full benefit of amplification. Prior to amplification, listeners with hearing impairment may experience difficulty perceiving subtle high-frequency changes in speech known as second-formant transitions. Second-formant transitions provide cues in discriminating consonant sounds. The purpose of this study was to determine if intensive high-frequency discrimination training for new hearing-aid users could facilitate improvements in speech perception. Speech recognition was evaluated for consonant identification, speech feature perception, and sentence identification in quiet and in noise before and after 10 days of training discrimination of frequency sweeps at 2 kHz. Improvements were found in sentence identification. These results have implications for the use of auditory training to improve speech recognition for newly amplified listeners with hearing impairment.

5aSCa16. Vowel production in Swedish children with cochlear implants. Tina Ibertsson, Birgitta Sahlen, and Anders Lofqvist (Dept. Logoped., Phoniatrics & Audiol., Clinical Sci., Lund, Univ. Hospital, SE-221 85 Lund, Sweden)

This study analyzed vowel production of Swedish children with cochlear implants. Swedish has a rich vowel system with nine vowels. Minimal word pairs were used as linguistic material. The age of the children varied between 6 and 13 years. The age at implantation was 1.2–10 years. The first two formants were measured for each vowel. To estimate the vowel space, the average Euclidean distance in the $F1$–$F2$ plane between each vowel and the mean first and second formant frequencies of all the vowels was calculated for each child. Preliminary results from 15 children with cochlear implants and 11 control subjects with normal hearing show no difference between the two groups, due to a large amount of variability in the group with cochlear implants. The relationship between the vowel space and the speech perception of each child with a cochlear implant was analyzed using correlations. Results show no strong correlation between the two measures, again due to the variability within the group. [Work supported by the Sunnderdahl Foundation.]

5aSCa17. Sentence duration and mode of communication in cochlear implanted children. Nicole L. Wiessner, Emily A. Tobey (Callier Adv. Hearing Res. Ctr., Univ. of Texas at Dallas, 1966 Inwood Rd., Dallas, TX 75235), and Ann E. Geers (Univ. of Texas Southwestern Medical Ctr. at Dallas, Dallas, TX 75390)

This study investigates the relationships between mode of communication and speech duration in cochlear implanted (CI) children with 4–6 years experience with their device. One hundred thirty-six CI children between the ages of 8 and 9 years old repeated sentences of three, five, and seven syllables (McGarr, 1983). Durations were examined according to syllable length, low versus high context sentences, and the communication

modality of the child's classroom. Findings suggest that communication modality is more strongly associated with duration than sentence context or length. Significantly shorter sentence durations were found for CI children taught with an auditory-oral emphasis than CI children taught using an emphasis of speech and some form of manual sign system. [Work supported by the NIH.]

5aSCa18. Interpolation of vocal-tract shape during stop closures from transition segments in vowel-consonant-vowel syllables. Milind S. Shah and Prem C. Pandey (Dept. of Elec. Eng., Indian Inst. of Technol., Bombay, Powai Mumbai 400076, India, pcpandey@ee.iitb.ac.in)

Children with prelingual profound hearing impairment have great difficulty in acquiring speech. Speech-training systems providing visual feedback of vocal-tract shape are found to be useful for improving vowel articulation. Vocal-tract shape estimation, based on LPC and other analysis techniques, generally fails during stop closures, and this restricts its effectiveness in speech training for production of consonants not having visible articulatory efforts. A technique based on two-dimensional surface modeling of the area values, estimated by LPC analysis, during the vowel-consonant and consonant-vowel transitions preceding and following the stop closure, has been investigated for interpolating the area values during the stop closures. Surface modeling was based on least-squares bivariate polynomials and Delaunay triangulation methods. Syllables of the type /aCa/, /aCi/, /iCa/, /iCi/, and /uCu/ with stop consonants /p/, /b/, /t/, /d/, /k/, and /g/ were analyzed for the estimation of place of stop closure. For bilabial, alveolar, and velar stops, the place could be estimated consistently with conic polynomial surface interpolation. Estimation of place, based on Delaunay surfaces, was consistent for bilabial and velar stops. However, cubic polynomial surface interpolation results were less consistent in estimating the place of constriction.

5aSCa19. Child voice and noise; acoustic effects of a day at the day-care related to background noise levels. Anita McAllister (INR, Speech Pathol., Linkoping Univ., Sweden), Svante Granqvist, Johan Sundberg, Peta Sjolander (Royal Inst of Technol.), and Mechtild Tronnier (Linkoping Univ.)

Several studies have found that high background noise levels are detrimental to health parameters. In particular, this seems to apply to developing voices were future vocal habits are established. Thus, it is important to study vocal function and environmental effects on the developing child voice. This study analyzed the effects of background noise on children's voices, specifically vocal intensity, and fundamental frequency. The investigated vocal parameters were (1) the relationship of background noise levels to F0 and vocal intensity, (2) F0 and vocal intensity variations over the day, and (3) F0 perturbation variations over the day. Ten 5-year-old children from three day-cares participated, six boys and four girls. The audio signal was recorded by two microphones mounted in front of the subjects' ears. By adding these signals it is possible to separate the voice from background noise. The material analyzed contained data from three 60-min recordings per child from morning, noon, and afternoon during a normal day at the day-care. Generally high mean background noise levels were found (82.6 dBA). Preliminary results suggest a correlation between high background noise and high F0 and vocal intensity in the children's voices, particularly for boys. F0 perturbation tends to increase during the day.

5aSCa20. Acoustic correlates of primary motor speech disorders in children during oral and hand tasks. Beate Peter and Carol Stoel-Gammon (Univ. of Washington, 1417 N.E. 42nd St., Seattle, WA 98195)

Primary child speech disorders have been subclassified variously. One proposed subtype, childhood apraxia of speech (CAS), results from motor programming deficits in speech production, and associated characteristics include limited phoneme inventory, variability in production, and diffi-

culty with speech prosody, e.g., monostress and monopitch. The purpose of this study was to identify acoustic correlates of oral and limb movements, derived from the hypothesis of a generalized timing deficit, that characterize this disorder subtype differentially from other disorder subtypes. Twelve children (age 4;7 to 6;6), of which eight were labeled with CAS, and 12 age- and gender-matched peers participated in a variety of oral motor and hand tasks. *K*-means clustering generated three disorder subgroups in the sample of participants with speech disorders, two of which were populated with CAS labels. Distinguishing characteristics included accuracy of vowel duration in a nonword imitation task, accuracy of clap interval duration in a clapped rhythm imitation task, rate of syllable repetition, durations of unstressed but not stressed vowel durations, and breath group duration. Taken together, these results are consistent with the view that deficits in timing accuracy, movement speed, and breath support underlie at least one subtype of idiopathic speech disorder.

5aSCa21. Changes in acoustic characteristics of /r/ following production training using electropatatography. Anna Schmidt (School of Speech Pathol. & Audiol., Kent State Univ., Kent, OH 44242)

Visual feedback using electropalatography (EPG) is an effective method for changing production of speech sounds in children. Differences in tongue placement and general acceptability of production have been demonstrated with EPG training for tongue to palate contact. Little is known, however, about changes in the spectral characteristics of speech sounds following training. The current study examined pre- and post-EPG patterns for /r/ in prevocalic, intervocalic, and postvocalic positions in sentences with 5 weeks of intensive EPG training. In particular, pre- and post-spectral characteristics of /r/ were examined for differences pre- and post-treatment. Subjects were four children (aged 9–10). These children exhibited distorted /r/ despite several years of traditional speech therapy but learned to produce perceptually acceptable /r/ with EPG. Spectral characteristics for /r/ productions in the same sentences by four typically developing children were used for comparison. Differences will be discussed in terms of acoustic characteristics as well and changes in tongue position.

5aSCa22. Physiologic and behavioral classification of delayed speech: Syllable repetition tasks. Jennell C. Vick, Christopher A. Moore, Lakshmi Venkatesh (Dept. of Speech & Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105, jennell@washington.edu), Thomas F. Campbell, Heather Leavy Rusiewicz (Univ. of Pittsburgh, Pittsburgh, PA), Lawrence D. Shriberg (Univ. of Wisconsin—Madison, Madison, WI), and Jordan R. Green (Univ. of Nebraska—Lincoln, Lincoln, NE)

As part of a large-scale study that aims to create a data-driven model of speech delay of unknown origin (SD) in preschool children, this investigation analyzed two- and three-syllable nonword repetitions by 50 3- to 5-year-old children, half of whom are typically developing (TD). The syllable repetition task was composed of four different words, bama, bada, bamana, and manaba, selected both for articulatory simplicity and to maximize displacement of the external articulators for kinematic analyses. While children in the TD group were significantly more accurate in both two- and three-syllable repetitions than children in the SD group, neither group was able to repeat three-syllable nonwords with particular acumen. Children in the TD group were 55% accurate, while children in the SD group were only 22% accurate. Measures of repetition accuracy, acoustics, nasalence, external articulatory kinematics, and respiration contribute to a hierarchical agglomerative clustering model that generates hypotheses about potential subclinical categories within the SD and TD groups.

5aSCa23. Changes of voice characteristics under noisy conditions: Acoustic and fiberscopic studies of school teachers. Noriko Kobayashi (School of Allied Health Sci., Kitasato Univ., 1-15-1, Kitasato, Sagami-hara, Kanagawa, 228-8555 Japan, noriko@ahs.kitasato-u.ac.jp) and Takashi Masaki (Kitasato Univ, Sagamihara, 228-8555 Japan)

It has been known that incidence of vocal pathology was high in school teachers as they had to use a great amount of voice and loud phonation in noisy school settings with the children's loud voices. Vocal hygiene programs and voice therapy are effective treatments for these cases. The final goal of the treatment should be to obtain the skills to use efficient and healthy phonation even in abusive and noisy environments. In our study, two school teachers with vocal nodules, two speech therapists who were trained to keep proper phonation even in noisy environments, and ten college students with no laryngeal pathology (control group) spoke eight sentences in two environments: a quiet environment and one with meaningful multitalker babble (MMB). Acoustic analyses and fiberscopic examination revealed higher sound pressure and F0 levels and laryngeal constriction in MMB for the teachers before therapy and the students, whereas two speech therapists kept similar F0 levels and the nonconstricted laryngeal condition in both environments. A teacher who had voice therapy produced less constricted voice with MMB than the pretherapy recording. The results suggested the efficacy of voice therapy with MMB for teachers to obtain efficient and healthy voice in their noisy work environments.

5aSCa24. Effect of speech task on intelligibility in speakers with dysarthria. Kate Bunton (Dept. of Speech, Lang., and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721-0071, bunton@u.arizona.edu)

There is some evidence that speech intelligibility varies across speech task for speakers with dysarthria. Primary differences have been related to the linguistic level of the material being evaluated (e.g., single words, sentences, conversation). While conversational speech is the most socially valid context for evaluating speech intelligibility, quantifying intelligibility in conversation limits its clinical use. The present study assessed intelligibility in speakers with dysarthria across four speech production tasks: single words, sentence production, passage reading, and spontaneous speech. Speakers varied with regard to the type of dysarthria, including hypokinetic, hyperkinetic, spastic, and mixed. Results show that all speakers were significantly less intelligible during spontaneous speech than the other structured speech production tasks (e.g., single words, sentences, and reading). Mean differences varied across speech severity level but did not vary across dysarthria type. These data indicate that speech production task is an important variable to consider during the evaluation of dysarthria and highlight the importance of developing techniques to measure the intelligibility of conversational speech. [Work supported by NIH Grant R03 DC005902.]

5aSCa25. Acoustic features of speech after glossectomy. Yuichi Ishimoto (School of Media Sci., Tokyo Univ. of Technol., 1404-1 Katakura, Hachioji, Tokyo 192-0982, Japan, ishi@media.teu.ac.jp) and Noriko Suzuki (Showa Univ., Ohtaku, Tokyo 145-8515, Japan)

The tongue serves an important role in speech production. Speech disorders due to tongue resection often occur and cause communication difficulties after surgery, according to the extent of the resected area. Such patients typically require appropriate speech therapy and palatal augmentation prosthesis. However, diagnosis of the speech in such patients has been based largely on subjective evaluation by clinicians. Acoustic features in glossectomee have yet to be clarified, and there is no objective estimation method for speech intelligibility. This study aims to clarify acoustic features associated with speech disorders after glossectomy and to develop a computerized evaluation system. The speech data were analyzed using the improved cepstrum method and compared with varying degrees of speech intelligibility. The vowel /i/ was selected as the speech sample because it is the high front vowel. The results showed that spectral

envelopes of the Japanese vowel /i/ with low speech intelligibility exhibit two peaks in the second formant (F2) region from 1500 to 3000 Hz. This suggests that the spectral features in the F2 region of the vowel /i/ are important clues to determine speech intelligibility.

5aSCa26. New generation aids for laryngectomy patients. Andrzej Czyzewski, Piotr Odya (Gdansk Univ. of Technol., Narutowicza 11/12, PL-80-952 Gdansk, Poland), and Bozena Kostek (Inst. of Physiol. and Pathol. of Hearing, Warsaw, Poland)

The artificial larynx has many disadvantages. The produced speech is monotonous and sounds artificially. In addition, produced speech intelligibility is usually poor. The major problem is a background noise caused by the device. In fact, the artificial larynx is only a simple vibrator, a construction of which has been almost unchanged since the 1950s. The aim of the presented project is to design a new generation of devices for laryngectomy patients. There are two different approaches to solve this task. The first one focuses on the artificial larynx. Some major improvements in the construction of the device might be easily introduced. Hence, the artificial larynx engineered was equipped with a digital processor and an amplifier. The spectral subtraction algorithm for noise reduction was used. In this method, an average signal spectrum and average noise spectrum are estimated and subtracted from each other, thus average signal-to-noise ratio (SNR) is improved. The main problem to be solved was that both the noise and the speech signal have the same excitation source and consequently are strongly correlated for voiced sounds. The second approach uses a PDA (portable digital assistant) to generate synthetic speech. The proposed new generation devices helping laryngectomy patients are presented in the paper.

5aSCa27. A speech communication aid system for total laryngectomies using voice conversion of body transmitted artificial speech. Keigo Nakamura, Tomoki Toda, Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Inst. of Sci. and Technol., 8916-5, Takayama-cho, Ikoma-shi, Nara, 630-0192 Japan, kei-naka@is.naist.jp)

There are several problems associated with using existing electrolarynxes. For example, the loud volume of the device itself might disturb smooth interpersonal communication, and its generated speech is also unnatural. To improve the quality of speech communication using such a medical device, this paper proposes a novel speech communication aid system for total laryngectomies. This system detects articulated speech caused by a new sound source as an alternative to the existing electrolarynx through the soft tissues of the head with a nonaudible murmur (NAM) microphone attached to the surface of the skin [Nakajima et al, Proc. Interspeech 2005, pp. 293–296 (2005)]. The new sound source outputs signals of extremely low energy that cannot be heard by people near the speaker. Such body-transmitted artificial speech is converted to a more natural voice by statistical voice conversion [T. Toda and K. Shikano, Proc. Interspeech 2005, pp. 1957–1960 (2005)]. The performance of the proposed system is evaluated in terms of objective and subjective measures using body-transmitted artificial speech simulated by a non-disabled speaker. Experimental results show that body-transmitted artificial speech is consistently converted to a much more natural and intelligible voice. [Work supported by SCOPE-S.]

5aSCa28. Development of a hands-free electrolarynx for persons whose larynx have been removed. Mitsuo Hashiba (Hokkaido Industrial Res. Inst., N19W11, Kita-ku, Sapporo, 060-0819, Japan), Yasunori Sugai (Densei Inc., Chuo-ku, Sapporo, 060-0002, Japan), Takashi Izumi (Hokkaido Tokai Univ., Minami-ku, Sapporo, 005-8601, Japan), Shuichi Ino, and Tohru Ifukube (The Univ. of Tokyo, Meguro-ku, Tokyo, 153-8904, Japan)

An electrolarynx is a useful speech-substitute device for patients who have lost laryngeal function. Unfortunately, the voices produced by a conventional electrolarynx are completely flat, unlike a human voice. For that reason, we developed a pitch-controlled electrolarynx that allows patients

to control a voice intonation using their exhalation [N. Uemi and T. Ifukube, "Design of a new electrolarynx having a pitch control function," IEEE RO-MAN'94, 198–203 (1994)]. This pitch-controlled electrolarynx has been manufactured and more than 3000 patients have used the device since 1998. However, many patients have strongly requested that we design a hands-free electrolarynx so that they can perform daily life functions and work in their office without their hands occupied by the electrolarynx control during use. In response to that request, we designed some models of a hands-free electrolarynx that is attachable to the patient's neck. Furthermore, by comparing the models based on evaluation tests regarding its sound quality and usability, a thermo-plastic brace was found to be the best for attachment to the neck. Although our hands-free electrolarynx has not yet been equipped with the pitch-control function, it was ascertained that it is useful for taking notes while talking on the phone, and so on.

5aSCa29. Modeling aspects of vocal fold oscillations with validation to clinical data. Sarah A. Bentil (Dept. of Mech. Eng., Univ. of Hawaii Manoa, Honolulu, HI 96822), Todd Reed, John S. AllenIII (Univ. of Hawaii Manoa, Honolulu, HI 96822), and Yuling Yan (Stanford Univ. School of Medicine, Stanford, CA 94305)

We investigate vocal fold vibrations in patients with muscle tension dysphonia (MTD) and Parkinson's diseases (PD) using a refined two mass model [I. Steinecke and H. Herzel, J. Acoust. Soc. Am. **97**(3), 1874–1884 (1995)] and clinical data obtained from acoustic recordings and high-speed digital imaging (HSDI) of the larynx. The two-mass model with empirical parameters is able to reproduce some of the behaviors revealed from functional, image-based analyses of vocal fold dynamics in both normal and abnormal states. In particular, the glottal area waveform (GAW) is used to describe the vocal fold vibratory behavior, which is derived from HSDI data using an analytical framework developed by the Yan research group [Yan et al., J. Voice **19**(2), 161–175 (2005); IEEE Trans. Biomed. Eng., (2006)]. Further, a comparison of the two-mass model predictions and the GAW and acoustic data was performed using signal processing techniques including fast fourier transforms, wavelet transforms, and spectrograms. These analyses deliver a comprehensive quantitative representation of the dynamic characteristics of the vocal folds and may provide new insights into the mechanism of abnormal voice productions such as in MTD and PD. [Research is sponsored by NSF Grant #0627992.]

5aSCa30. Preshearing effects on viscoelastic properties of collagen-based vocal fold injectables. Sarah A. Klemuk (Dept. of Speech Pathol. & Audiol., Univ. of Iowa, 336C SHC, Iowa City, IA 52242, sarah-klemuk@uiowa.edu) and Ingo R. Titze (Univ. of Iowa, Iowa City, IA 52242)

A criterion for developing a biomaterial is that its viscoelastic properties correspond to native tissues. This criterion is particularly important for vocal fold tissues. Collagen-based injectables are routinely used to repair paralyzed or scarred vocal folds and restore vocalization, yet reported viscoelastic properties are inconsistent. The present study evaluated the effects of preshearing samples on their viscoelastic measurements. Non-cross-linked bovine collagen (Zyderm II[TM]) and micronized dermal tissue (Cymetra[TM]) were presheared by extruding directly through a mechanically operated syringe with different needle sizes. Preshear rates ranged from 31 to 5600/s for orifice radii of 1.21 to 0.0955 mm, respectively. Sandpaper-coated, parallel plate attachments on a stress-controlled rheometer were used to measure elastic moduli and dynamic viscosities from 0.01 to 10 Hz. A 3–10-fold reduction in elastic modulus and a 2–4-fold reduction in dynamic viscosity occurred for preshear rates exceeding 1000/s compared to rates of 31 and 131/s. These results suggest a change in the molecular structure and the viscoelastic compatibility of vocal fold injectables when sufficiently presheared. [Work supported by NIH.]

5aSCa31. Spasmodic dysphonia in Japan: 1486 Botulinum toxin injections for 260 spasmodic dysphonia patients. Masanobu Kumada (Kumada Ear-Nose-and-Throat Clinic, L-First Bldg. 3F, 4-2-6 Nishi-Azabu, Minato-Ku, Tokyo, 106-0031 Japan) and Takeo Kobayashi (Teikyo Univ. Ichihara Hospital)

Botulinum toxin injection is a very effective treatment for spasmodic dysphonia. We experienced 1486 injections for 260 patients in past 15 years in Japan. Here we show statistics of these injections. We use type A toxin that is produced by Clostridium Botulinum. This toxin affects the neuro-muscular junction as a blocker of release of Ach, which leads to inversible parasis of a muscle. Mainly, we inject the toxin paracutaneously on the neck to the vocalis muscle using a partly covered needle that works as an EMG electrode. Mainly we have four options: 2.5 i.u. unilaterally (38% of 256 patients), 2.5 i.u. for each laterality bilaterally (16%), 5.0 i.u. unilaterally (29%), and 5.0 i.u. for each laterality bilaterally (5%). Its effectiveness persisits for 17.3 weeks on average. The longest effectiveness we have ever experienced is more than 1 year. A small percentage of the injections are not effective; it can be thought that the injected toxin does not reach the neuromuscular junction even when the needle is in the muscle: neuromuscular junctions contribute a part of the muscle and EMG signals do not tell us where these junctions are. Mainly, the side effects are breathy hoarseness and misdiglutition, which persist a week or so.

5aSCa32. Collagen injection as an adjunct to arytenoid adduction. Miwako Kimura, Takaharu Nito (Dept. of Otorhinolaryngology, The Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 1138655, Japan, mkimu-tky@umin.ac.jp), Niro Tayama (Intl. Medical Ctr., Shinjuku-ku, 162-8655, Tokyo, Japan), and Roger Chan (Texas Univ., TX)

Dysphonia associated with laryngeal paralysis may be identified in the short term postoperatively or may develop years after successful arytenoid adduction. In selected cases, collagen injection of vocal fold improves phonation after arytenoid adduction. Our experience with the use of collagen injection to supplement arytenoid adduction is reported. Forty patients were treated with transoral collagen injection into the atrophic vocal fold after arytenoid adduction at The University of Tokyo from January 1990 to December 2005. These patients all had a diagnosis of unilateral vocal fold paralysis. Voice quality (GRBAS scale) and the other examinations for vocal function demonstrated measurable improvement after treatment. After arytenoid adduction, maximum phonation time (MPT) improved from 5 to 11 s and mean flow rate (MFR) as an aerodynamic examination improved from 772 to 322 ml/s. Voice quality, especially roughness, improved from 1.8 to 0.8 after collagen injection. MPT (from 11 to 14 s) and MFR (from 322 to 233 ml/s) showed overall improvement after collagen injection. We will report the ease with which transoral collagen injection was accomplished and how better voice quality was attained without the more extensive surgery. In the session techniques of anesthesia, injection, and patient selection are discussed.

5aSCa33. Numerical simulation of sound originated from the vocal tract in soft neck tissues. Makoto Otani, Tatsuya Hirahara (Dept. of Intelligent Systems Design Eng., Toyama PRef. Univ., Imizu, Toyama 939-0398, Japan, otani@pu-toyama.ac.jp), and Seiji Adachi (Fraunhofer Inst. for Bldg. Phys., 70569 Stuttgart, Germany)

A nonaudible murmur (NAM), a very weak speech sound produced without vocal vibration, can be detected by a special NAM microphone attached to the neck, thereby providing a new communication tool for use with functional speech disorders. The microphone is a condenser microphone covered with soft-silicone impression material that provides good impedance matching with soft tissues that are found in the neck. The NAM detected with the NAM microphone, however, is insufficiently clear because higher-frequency components are suppressed severely. A production mechanism of the NAM as well as transfer characteristics of the NAM in soft tissues of a neck should be clarified to improve the NAM clarity. As a first step, sound propagation from the vocal tract to the neck surface is simulated using a head model scanned with magnetic resonance imaging and finite difference method. Results show that soft-meat-tissue-conducted sound has fewer higher-frequency components than an air-conducted sound. The decay of transfer characteristics is approximately -10 dB/oct in the audible range, which is roughly parallel to the spectral characteristics of the measured NAM. [Work supported by SCOPE-S.]

Session 5aSCb

Speech Communication: Production I (Poster Session)

Dani M. Byrd, Cochair
Univ. of Southern California, Linguistics, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693

Kiyoshi Honda, Cochair
ATR Human Information Processing Labs., 2-2 Hikaridai, Seika-Cho Soraku-Gun, Kyoto 619-0288, Japan

Contributed Papers

All posters will be on display and all contributors will be at their posters from 11:00 a.m. to 12:00 noon.

5aSCb1. Computer graphics facial animation of Japanese speech using three-dimensional dynamic visemes. Makoto J. Hirayama (Kanazawa Inst. of Technol., 7-1 Ohgigaoka, Nonoichi Ishikawa, 921-8501 Japan, mako@infor.kanazawa-it.ac.jp)

Speech production makes not only acoustic signals but also visual images of the face especially for the jaw, lips, teeth, and tongue. To reproduce realistic facial images during Japanese speech, a three-dimensional computer graphics model of Japanese visemes was made. These visemes were extracted from the video database of images around lips during speech captured by two high speed (up to 300 frames per second) cameras [M. J. Hirayama, Proc. ICPhS 2003 (2003), pp. 3157–3161]. Most of the visemes are created as static shapes. They are for five vowels, semi-vowels, and some consonants. For explosives by labials (/p/ /b/ /n/) or tongue (/t/ /d/ /n/), dynamic information, that is, multiple shapes and timing information were assigned for each viseme. By placing these visemes onto a time axis at the timing of phonemes of a sentence, then interpolating shapes in between by using a spline interpolation technique on speech articulators' motion graphs, computer graphics animation was made by ray-tracing rendering. [A part of this work was supported by Japan MEXT Academic Frontier Project.]

5aSCb2. A photoglottographic (PGG) method with external lighting and sensing. Kiyoshi Honda (LPP-UMR7019, CNRS/Univ. Paris 3, Paris F75005 France, and ATR, Kyoto 619-0288, Japan, honda@atr.jp), Shinji Maeda (CNRS/ENST, Paris, F75634 Cedex13 Paris, France), Stephane Hans (Service ORL, F75015 Paris, France), and Lise Crevier-Buchman (CNRS/Univ. Paris 3, Paris F75005 France)

Photoglottography (PGG) is a moderately invasive technique to observe glottal movements and vocal-fold vibrations. A light guide is inserted into the pharynx and transglottal light is detected by a photosensor placed on the outer skin below the glottis, or conversely the sensor is placed inside the pharynx and the subglottal cavity is illuminated from outside with a light source. This report evaluates a noninvasive PGG method, in which the hypopharynx is illuminated with a light guide placed on the neck skin at the level between the hyoid bone and thyroid cartilage and a photosensor on the skin near the cricoid cartilage. The light passed through the skin illuminates the lumen of the hypopharynx and the photosensor detects indirect (diffusion) light through the glottis during glottal adduction. Choosing an appropriate location for the light source and photosensor, glottal adduction pattern is observed for voiceless stops and fricatives in low as well as high vowel context. Thus, our diffusion-light

PGG potentially overcomes the tongue-position dependency of the detected signals, which has been a major caveat of the conventional PGG methods.

5aSCb3. Evaluation of a photonasography. Angélique Amelot (LPP, 19 rue des Bernardins 75005 Paris, angelique.amelot@univ-paris3.fr), Kiyoshi Honda, Shinji Maeda (ENST, 75013 Paris, France), Lise Crevier-Buchman (HEGP, 75015 Paris, France), and Patricia Basset (LPP, 75005 Paris, France)

A photonasographic (PNG) technique was tested to evaluate velopharyngeal port (VP) opening/closing during speech. This system uses a photodetector placed in the pharynx via a nostril and a fiberscope in another nostril pointing to the back wall of the nasopharynx. With this setting, the light is reflected from the walls, passes through the open port, and is sensed by the photodetector. In a pilot experiment, a native French speaker read sequences C1V1C1 V2C1V1, where C1=/n/ or /t/ and V1,2=/a/, /i/, /A/, /E/, or /O/, as /tatA ta/. The obtained PNG data compare qualitatively well with the velum position data, on the same corpus, derived from the naso-fiberscopic video images. When NPG detects a VP opening, the fiberscope data correspondingly indicate a lowering of the velum and inversely. PNG, as well as the velum position data, show that VP is open at the onset of a nasal phoneme and is closed just after the offset of that nasal phoneme. Moreover, VP is not completely closed during the first and second phoneme in the sequence [ninini] and [nanana], indicating their nasalization. The combination of a well-set fiberscopic light-source and photodetector appears to give a good estimation of the degrees of VP opening.

5aSCb4. An MRI-based 3-D tongue atlas for tongue modeling. Chao-Min Wu and Han Tiet Goh (Dept. of Elec. Eng., Natl. Central Univ., 300, Chun-Da Rd., Chung-Li, Taiwan 320)

The main objective of this study is to build an MRI-based 3-D tongue atlas with an established available spatial transformation technique. The subjects for the MRI data were eight male and eight female college students (19–28 years old) who are native speakers of mandarin with Taiwanese accent without speech disorders. The oral MR images (axial: TR, 400 ms; TE, 10 ms; FOV 24×24; image matrix, 256×256 for 35 slices with 2-mm thickness) were acquired using a GE SIGNA 1.5-T scanner in the University Hospital of Chung Shan Medical University. The axial MR images of the human tongue were first segmented with snakes active contour method, then the 3-D tongues of each subject were reconstructed with morphology-based gray-level interpolation. Finally, these 3-D tongues were spatial transformed into a 3-D tongue atlas with thin-plate spline method. The 3-D tongue atlases for male and female subjects and mor-

phometric studies, such as subject-to-subject, subject-to-atlas, and male-to-female, of these two atlases will be presented. Implication of these data related to tongue modeling will also be discussed. These information will be used as the basis for future research on tongue modeling and underlying physiological mechanism of normal and disordered speech production.

5aSCb5. Direct measurement of glottal volume velocity using high-speed, stereoscopic, particle imaging velocimetry. David A. Berry, Juergen Neubauer, and Zhaoyan Zhang (The Laryngeal Dynam. Lab., UCLA Div. of Head and Neck Surgery, 1000 Veteran Ave., Rm. 31-24, Los Angeles, CA 90095-1794)

Despite its central importance as the glottal source in several long-standing models of voice production, the glottal volume velocity signal has been relatively little studied. In particular, few direct measurements of the glottal volume velocity signal have been made. One common method for estimating the glottal volume velocity signal has been linear inverse filtering of the oral acoustic output. However, a major assumption of linear inverse filtering, and of the linear source-filter theory of speech production in general, is that the glottal source and the vocal tract are independent of each other. The accuracy of this assumption has been questioned, particularly at extremes of fundamental frequency and intensity. In order to test the limits of linear inverse filtering, direct, instantaneous measurements of glottal volume velocity have been made using high-speed, stereoscopic, particle imaging velocity on a self-oscillating, physical model of vocal-fold vibration. These direct measurements are contrasted with results from linear inverse filtering.

5aSCb6. Lip protrusion and tongue position in French vowels produced by blind speakers and sighted speakers. Lucie Menard, Annie Leclerc, Jerome Aubin, and Annie Brasseur (Phonet. Lab., UQAM, Case Postale 8888, succ. Ctr.-Ville, Montreal, H3C 3P8, Canada)

The influence of visual experience in speech production was investigated through a study of the French vowels /i y u a/. Articulatory and acoustic recordings of ten repetitions of the vowels in CVC syllables embedded in carrier sentences were conducted in three consonant environments (/b d g/) and two prosodic conditions (neutral and under contrastive emphasis). Six congenitally blind adults and six sighted adults were recorded. All participants were native speakers of Canadian French and had no history of speech disorder. The audio signal, lip movement, and tongue shapes were recorded using a digital camera and an ultrasound system. The minimal (for /i a/) or maximal (for /y u/) horizontal position of the upper lip was tracked and the front-back position of the tongue was measured for each vowel. Formant frequencies were extracted using LPC analyses. Results show that, despite similar acoustic differences, the rounded and unrounded vowels are less differentiated along the protrusion dimension for blind speakers compared to sighted speakers. It is suggested that this variability is related to a trade-off between lip protrusion and tongue position. Significant interaction effects between consonant environment and vowel category are discussed for each speaker group. [Work supported by SSHRC, NSERC.]

5aSCb7. Tongue muscle fiber tracking during tongue protrusion and rest. Hideo Shinagawa, Emi Z. Murano (Dept. of Biomed. Sci., Univ. of Maryland Dental School, 666 W. Baltimore St., Baltimore, MD 21201, hshinagawa@umaryland.edu), Jiachen Zhuo, Rao P. Gullapalli (Univ. of Maryland Medical Ctr., Baltimore, MD 21201), Bennett Landman, Jerry L. Prince (Johns Hopkins Univ., Baltimore, MD 21218), and Maureen Stone (Univ. of Maryland Dental School, Baltimore, MD 21201)

Diffusion tensor imaging (DTI) has recently become a promising tool to investigate not only the nerve structure of the brain, but also integration of the peripheral muscle structure. The purpose of this study was to represent the functional deformation of tongue muscles with DTI by exam-

ining two conditions, tongue protrusion and rest. Three normal volunteers participated in the study. They were recorded at rest and wearing a custom made tongue retaining device (TRD), often used for patients with sleep apnea. The TRD held the tongue immobile in protruded position. In both positions, the tongue muscle structure and functional deformation could be visualized. In particular, the deformation of the genioglossus (GG) muscle, which is generally fan shaped, was clearly visible. The fibers of GG anterior muscle were more vertical with the TRD. The DTI technique has the potential to reveal not only myoarchitecture but also functional deformation of the tongue *in vivo*. With a little ingenuity, it might be possible to conquer the MRI sensitivity to small motion artifacts or improve tongue muscle stabilization to track the muscle fiber architecture during sustained speech. [Work supported by JSPS and NIH.]

5aSCb8. The effects of the false vocal folds on laryngeal pressures and flows. Li Sheng (Dept. of Biomed. Eng., School of Life Sci. and Technol., Xi'an Jiaotong Univ., Xi'an 710049), Ronald C. Scherer (Bowling Green State Univ., Bowling Green, OH 43403), Wan MingXi, Wang SuPin, and Qi LiYun (Xi'an Jiaotong Univ., Xi'an 710049)

As a laryngeal constriction, the false vocal fold (FVF) gap (the distance between the medial edges of the FVFs) may create important effects on phonation by altering the pressures within and flows through the larynx. The computational code FLUENT was used to examine the effects on pressures and flows for FVF gaps ranging from 0.02 to 2.06 cm, for three glottal angles (uniform and convergent/divergent 40 deg) and two minimal glottal diameters (0.04 and 0.06 cm), for constant subglottal pressure (8 cm H_2O). The specific design of the FVFs followed prior literature. Results suggest three important ranges of FVF gaps: (1) when the FVF gap was 1.5–2 times the minimal glottal diameter, pressures were lower throughout the larynx, and flows were higher through the larynx (less flow resistance); (2) for smaller FVF gaps, intralaryngeal pressures increased and flows decreased; (3) for greater FVF gaps, pressures and flows were unaffected. Also, (4) the divergent glottal angle gave the greatest flows, (5) flow separation locations for the divergent glottis moved downstream as the FVF gap decreased, and (6) the FVFs straightened the flow for narrow gaps. Thus, the status of the FVFs needs to be considered for voice production and synthesis.

5aSCb9. Acoustic and aerodynamic effects of false folds and epiglottis. Fariborz Alipour, Sanyukta Jaiswal, and Eileen Finnegan (Dept. Speech Path. and Audiol., Univ. of Iowa, Iowa City, IA 52242)

Excised larynx research usually excludes the false folds in an effort to reveal the true vocal folds. The purpose of this study was to examine the aerodynamic and acoustic effects of the false folds and epiglottis on excised larynx phonation. Canine larynges were mounted over a tapered tube on the excised larynx bench that supplied pressurized, heated, and humidified air. Glottal adduction was accomplished by passing sutures that simulated lateral cricoarytenoid muscle activation. First, the excised larynx with intact false folds and epiglottis was subjected to a series of pressure-flow experiments with tension and adduction as major control parameters. Then, the epiglottis and false folds were successively removed and the experiment was repeated for each condition. The subglottal pressure, EGG, flow rate, audio signal, and sound pressure level were recorded during each experiment. Glottal flow resistance was calculated from the pressure and flow signals, while the EGG signal was used to extract fundamental frequency. It was found that these additional structures have positive contribution to the glottal resistance and sound intensity of the larynx. Also, vocal fold elongation and glottal medial compression caused an increase in the glottal resistance. [Work supported by NIDCD Grant No. DC03566.]

Session 5aUW

Underwater Acoustics and Acoustical Oceanography: Geoacoustic Inversion

Dezhang Chu, Cochair

Woods Hole Oceanographic Inst., Dept. of Applied Physics and Engineering, Woods Hole, MA 02543

Hiroyuki Hachiya, Cochair

Chiba Univ., Research Ctr. for Frontier Medical Engineering, 1-33 Yayoi-Cho, Inake-ku, Chiba 263-8522, Japan

Chair's Introduction—7:45

Contributed Papers

7:50

5aUW1. Notes on broadband seabed geoacoustic inversion in shallow water. Ji-Xun Zhou and Xue-Zhen Zhang (Georgia Inst. of Technol., Atlanta, GA 30332-0405 and Inst. of Acoust., Chinese Acad. of Sci., Beijing, 100080, China)

With increasing interest in shallow-water environments, where bottom interaction is a dominant effect, there is a need to acquire accurate information about seabed geoacoustic parameters. Due to the difficulties and high costs of directly measuring seabed parameters at low frequencies (LF) in shallow water (SW), seabed geoacoustic inversion from long-range propagation has become an active research area. Much progress on this subject, notably on inversion methods, has been made in the last two decades, mainly through the use of powerful numerical codes and data processing tools. Despite this, seabed geoacoustic inversion often fails to yield systematic LF seabed parameters in a broad band (such as speed/attenuation vs frequency) that are required for seabed geoacoustic modeling or for sonar performance prediction in the SW environment. This paper will discuss several basic concepts and technical issues related to SW geoacoustic inversion, including filter phase shifting, speed-attenuation coupling, hidden depth, and sea-surface effects. It shows that inverting the broadband seabed acoustic parameters is a delicate task that can often be subject to errors. These issues should be considered in designing experiments that test seabed geoacoustic inversion methods or in interpreting experimental results. [Work supported by ONR and NNSF of China.]

8:05

5aUW2. Efficient use of *a priori* data in sediment inversions through the use of null space. Gopu R. Potty, James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882), Ying-Tsong Lin, and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA)

We present improved inversions for sediment parameters by incorporating *a priori* information about the environment. This improvement is achieved by projecting a desirable solution into the null space of the inversion and including this null space contribution along with the standard non-null space contribution. We use singular value decomposition (SVD) to define the null space of the inversion and elucidate our projection method. The desirable solution, which is projected into the null space, is constructed based on previous data from cores, geophysical surveys, and historic data. This approach introduces user bias into the solution; the projection onto the null space supplies a safety net by showing which aspects of this bias are justified by the data. The user bias can be considered as additional data which, when incorporated, can lead to meaningful solutions. Effectiveness of probing the null space will be compared to the more conventional nonlinear inversion schemes. This approach will be tested using field data collected as part of the Shelfbreak Primer experiment. [Work supported by ONR.]

8:20

5aUW3. Robust source localization and geoacoustic inversion in the Haro Strait Primer. Rashi Jain, Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102), and Alex Tolstoy (A. Tolstoy Sci., McLean, VA 22101)

Gibbs sampling, a Markov chain Monte Carlo technique, has been shown to be a powerful tool for geoacoustic inversion and source localization. By providing estimates of posterior joint distributions, it offers a global optimization route for multidimensional estimation that reports uncertainty and covariance in addition to point estimates. In this work, Gibbs sampling is applied for extracting time delays from recorded time series during the Haro Strait primer experiment. Employing time delay estimates and using a linear approximation to the inverse problem and then regularization, estimates are obtained for source and receiver location and some environmental parameters. Simultaneously processing receptions at all three vertical line arrays for localization of each source reduces ambiguities in the estimation process. Similarly, using received signals corresponding to several sources, array element localization for a single array, a difficult problem for this particular data set, becomes more precise. Multiple data sets are used in the inversion and consistent results validate the robustness of the approach. Estimated bathymetry is in agreement with bathymetric maps for the region. [Work supported by ONR.]

8:35

5aUW4. Coherent noise processing and geoacoustic inversion. Peter Gerstoft, Chen-Fen Huang, and William S. Hodgkiss (Marine Physical Lab., Univ. of California San Diego, San Diego, CA 92093-0238)

Ocean acoustic noise can efficiently be processed to extract Green function information from noise [Roux *et al.*, J. Acoust. Soc. Am. (2004), Siderius *et al.*, *ibid.* (2006)]. By cross-correlating the ambient noise field from two sensors, it is possible to extract the impulse response between the two sensors including bottom and subbottom bounces. When this noise processing is used on a vertical array, it can give valuable information about the subbottom near the array. This information will be then used to constrain a classical geoacoustic inversion procedure where we use a distant towed source to obtain the geoacoustic bottom parameters.

8:50

5aUW5. Geoacoustic inversion based on both acoustic pressure and particle velocity. A. Vincent van Leijen (NLDA, P.O. Box 10.000, 1780 CA, Den Helder, The Netherlands, av.vanleijen@kim.nl), Jean-Pierre Hermand (ULB, B-1050 Bruxelles, Belgium), and Kevin B. Smith (NPS, Monterey, CA 93943)

Conventional inversion schemes for environmental assessment depend on an objective function that exploits amplitude or phase information of acoustic pressure data alone. This work investigates the potential of vector sensors for geoacoustic inversion by defining an objective function that

5a SAT. AM

3355 J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ **3355**

also takes into account acoustic particle motion. Calculations are performed on synthetic broadband data for a shallow water environment (South Elba) with an optimization scheme based on different metaheuristics. Differences in the inversion process, including sensitivity of the cost function to environmental parameters and convergence speed of the optimization algorithm, are presented by comparing inversion results for a sparse pressure-only array and a vector sensor array.

9:05

5aUW6. Geophysical parameter inversion in a range-dependent environment. Woojae Seong, Keunhwa Lee, Kyungsup Kim (Dept. of Ocean Eng., Seoul Natl. Univ., Seoul, 151-742, Korea), and Seongil Kim (Agency for Defense Development, Jinhae, Korea)

Matched-field inversion technique is applied for estimation of geophysical parameters of the ocean bottom in a range-dependent shallow water. In the experiment (MAPLE-4), conducted off the coast of the East Sea during May 2005, narrow-band multitone cw acoustic data were obtained from the towed moving source along a weakly range-dependent track, from 2 to 18 km apart from the L-shaped receiver array. In the inversion, complex density model based on Biot model is used to invert for parameters including porosity and permeability. Inversion results are compared with existing geological survey data. In addition, the effect of range dependency resulting from the seafloor slope and the existing bottom intrusion is examined.

9:20

5aUW7. Bayesian inversion of propagation and reverberation data. Peter L. Nielsen (NATO Undersea Res. Ctr., Viale S. Bartolomeo 400, 19138 La Spezia, Italy) and Stan E. Dosso (Univ. of Victoria, Victoria, BC, Canada V8W 3P6)

A Bayesian matched-field inversion approach to infer geoacoustic and scattering properties of the seabed is applied to simulated propagation and reverberation data received on a towed horizontal array. The approach is based on the method of fast Gibbs sampling (FGS) of the posterior probability density to estimate uncertainties in both geoacoustic and scattering parameters for broadband acoustic data in realistic shallow-water environments. The FGS is linked to an acoustic propagation model that simultaneously provides complex acoustic pressure at short propagation ranges and long-range reverberation intensity. The inversion algorithm is initially applied to long-range reverberation data alone to assess the geoacoustic information content of reverberation in terms of marginal posteriori probability densities for the environmental parameters. A reduction in uncertainty for the extracted geoacoustic and scattering parameters is demonstrated by a simultaneous inversion of the propagation and reverberation horizontal array data.

9:35

5aUW8. On the use of acoustic particle velocity fields in adjoint-based inversion. Matthias Meyer, Jean-Pierre Hermand (Université libre de Bruxelles, Belgium & Royal Netherlands Naval College, The Netherlands), and Kevin B. Smith (Naval Postgraduate School, Monterey, CA)

Following the recent interest in the use of combined pressure and particle motion sensors in underwater acoustics and signal processing, some general aspects regarding the modeling and multipath phenomenology of acoustic particle velocity fields in shallow water environments have been studied. In this paper we will address a number of issues associated with the incorporation of vector sensor data (pressure and particle velocity) into adjoint-based inversion schemes. Specifically, we will discuss the ability of a semi-automatic adjoint approach to compute the necessary gradient information without the need for an analytic model of the adjoint particle velocity field. Solutions to the forward propagation of acoustic pressure are computed using an implicit finite-difference parabolic equation solver while the particle velocity is calculated locally at each grid point. Some numerical examples of vector sensor inversion results are provided. [Work supported by Royal Netherlands Navy.]

9:50

5aUW9. Sensitivity analysis of a geoacoustic parametrized model and its application to inversion of seabed properties. Jin-Yuan Liu and Chung-Ray Chu (No. 70 Lien-hai Rd. Kaohsiung 804, Taiwan)

The work first aims to analyze the parametrized geoacoustic model proposed by Robins [J. Acoust. Soc. Am. **89**, 1686–1696 (1991)], in which the density and sound speed distributions vary with respect to depth as a generalized-exponential and an inverse-square function, respectively. The model contains a set of parameters that, by appropriate selection, may fit well the realistic geoacoustic variations [E. L. Hamilton, J. Acoust. Soc. Am. **68**, 1313–1340 (1990)]. By choosing the plane-wave reflection field as an objective function, each model parameter is carefully analyzed to determine its range and sensitivity. Then, numerical simulation is employed to establish an inversion procedure, in conjunction with the application of acoustic wave reflection from a nonuniform seabed. Finally, a field experiment is designed and implemented to estimate the seabed acoustic properties based upon the model parameter inversion.

10:05–10:20   Break

10:20

5aUW10. Characterization of sediment dynamics in an estuary environment using acoustic techniques. Jean-Pierre Hermand, Laura Perichon (Environ. Hydroacoustics Lab., Optics and Acoust. Dept., ULB-CP 194/05, 50 AV. F.D. Roosevelt, B-1050 Brussels, Belgium), and Michel Verbanck (Univ. Libre de Bruxelles, B-1050 Brussels, Belgium)

In recent years, acoustic-based methods have been developed to characterize the dynamical behavior of loose sediments and bed deposits in very shallow water environments. In this paper, we present preliminary results on the estimation of the dynamic changes in an estuarine environment using data from dual-frequency echosounding at high resolution and contemporaneous hydrological measurements including suspended matter concentration, density subbottom profiling, and data assimilation based on a sediment transport model. Those measurements are being conducted in the lower estuary of the Scheldt (Belgium) at the Sint Anna site where strong tide and season-dependent phenomena can be observed. This allows us to construct a ground-truthed, time-dependent geoacoustic model of the environment, i.e., a characterization of sound speed, density, and attenuation in function of time and depth. Synthetic acoustic data generated by that model will then be used to test inversion methods for monitoring sediment dynamics in real time.

10:35

5aUW11. Geoacoustic model for the New Jersey Shelf by inverting airgun data. Yong-Min Jiang, N. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 3055, Victoria, BC V8W 3P6, Canada), and Mohsen Badiey (Univ. of Delaware, Newark, NJ 19716)

This paper describes geoacoustic inversion of airgun data acquired during the SWARM95 experiment. Hybrid optimization and Bayesian inversion techniques were applied to three airgun data sets recorded by a vertical line array. Optimization results are used to show the consistency of the estimates from all of the shots in terms of histograms and standard deviations of the inverted geoacoustic model parameters. The inversion results from the Bayesian approach are used to show the uncertainties of the estimates in terms of marginal distributions, MAP estimates, and credibility intervals. In the Bayesian inversion, full data error covariance matrices were estimated by ensemble averaging the covariance of the residuals of the measured and modeled data of inversions from many shots. The numbers of shots in the ensemble averages were determined by checking the temporal coherence of the signal. Statistical tests were used to test the validity of the assumptions in the Bayesian approach after incorporating full data error covariance matrices. With these inversion techniques, equivalent geoacoustic models with/without shear wave estimates are extracted for this experimental site. The frequency dependence of the p-wave attenuation, and the correlation between the geoacoustic parameters are obtained from the inversion results. [Work supported by ONR.]

5aUW12. Mismatch effects on passive matched-field, thick-sediment inversion. Donald R. DelBalzo, Robert S. Linzell, and Lisa A. Pflug (Planning Systems Inc., 40201 Hwy. 190 E, Slidell, LA 70461)

Acoustic field predictions in shallow-water areas can be severely limited by environmental uncertainty, especially in sediment property characteristics. Direct methods of obtaining geophysical data are expensive, so inverse methods, which invert signals from controlled active sources, have recently become popular. This work assesses the impact of environmental mismatch on low-frequency (<100 Hz) matched-field correlations on broadband signals from surface ships with unknown source levels at unknown ranges. A range-staggered technique is employed to create a range-dependent sediment description, with ever-increasing confidence as additional ships are analyzed. Matched-phase techniques are applied in a simulated shallow-water environment with a single vertical array and high signal-to-noise ratios. The simulations indicate significant potential for passive, range-dependent, thick-sediment characterizations out to ranges of tens of water depths in shallow water, despite reasonable mismatch conditions in the water and in the sediments. [Work sponsored by SPAWAR.]

11:05

5aUW13. To obtain geo-acoustic parameters by the acoustic inversion technique using subbottom profiler. Chung-Wu Wang and Cheng-Che Lee (China College of Marine Technol. and Commerce, No. 212, Sec. 9, Yan-Ping N. Rd.,Taipei, Taiwan, ROC)

The fast sound speed at sea surface in the water surrounding Taiwan causes strong refraction. If the water depth is shallower than 5000 m, the sound ray will reflect at the bottom, causing severe bottom losses. Therefore, geo-acoustic parameters are important for acoustic studies in the water surrounding Taiwan. In the past, researchers used single-pulse sonar of 3.5 kHz to obtain geo-acoustic parameters by the spectral ratio technique. Now the sediment exploration has already advanced and used the Chirp Sonar system on Taiwanese Research Vessels such as Ocean Researcher I (OR I) and Ta-Kuans TOPAS system. This paper presents the results of attenuation coefficients in ocean bottoms using the acoustic inversion technique on the chirp sonar data collected by OR I. The results are compared with the core data, and the error is within 15%.

11:20

5aUW14. The resolution of a nonlinear ocean acoustic inverse problem. Robert I. Odom and Andrew A. Ganse (Appl. Phys. Lab. and Dept. of Earth and Space Sci., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698)

For a strictly linear inverse problem the model resolution matrix is a simple product of the linear system matrix and its generalized inverse. How much the model resolution differs from the identity is an indication of bias in the inverted model. Nonlinear inverse problems lead to higher order resolution operators, which allow one to quantify the effect of ignoring the nonlinearity in the inverse problem. In addition, the trade-off between resolution and variance is examined for the nonlinear inverse problem. Ocean acoustic inverse problems for ocean and/or bottom structure are intrinsically underdetermined, because some continuous quantity such as sound speed or temperature is estimated from a finite number of data values, and the model is generally discretely parametrized to be numerically tractable. Discretizing the model has the positive effect of regularizing an intrinsically ill-posed problem. However, discretizing an intrinsically nonlinear problem extracts a penalty. Higher order resolution operators may contribute even when they should be identically zero. (For example, in a quadratically nonlinear model, cubic and higher order resolution operators should be zero, but won't be zero as a result of the discretization.) Modal propagation in an ocean acoustic waveguide is used to illustrate the issues. [Work supported by ONR.]

11:35

5aUW15. Adapting results in filtering theory to inverse theory, to address the statistics of nonlinear geoacoustic inverse problems. Andrew A. Ganse and Robert I. Odom (Appl. Phys. Lab. and Dept of Earth and Space Sci., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

The intrinsically non-Gaussian statistics of nonlinear inverse problems, including ocean geoacoustic problems, is explored via analytic rather than numerical means. While Monte Carlo Bayesian methods do address the non-Gaussian statistics in nonlinear inverse problems, they can be very slow, and intuitive interpretation of the results are at times problematic. There is great theoretical overlap between recursive filters/smoothers, such as the extended Kalman filter, and methods of linear and nonlinear geophysical inversion. The use of recursive filters in inversion is not in itself new, but our interest is in adapting statistical developments from one to the other. Classic analytic methods in both filtering theory and inverse theory assume Gaussian probability distributions, but newer nonlinear filters do not all make this assumption and are explored for their potential application to nonlinear inverse problems. The similarities and differences between the frameworks of filtering theory and inverse theory are laid out in a series of geoacoustic inverse problem examples. Recent work in nonlinear filters handles non-Gaussian probability densities that are constrained to a particular form, and also derives analytic expressions for higher order moments of these density functions. The application of these developments to geoacoustic inverse problems is addressed. [Work supported by ONR.]

Session 5pAA

Architectural Acoustics, Speech Communication and Psychological and Physiological Acoustics: Psychological Aspects of Speech in Rooms II

John S. Bradley, Cochair

National Research Council, Inst. for Research in Construction, Ottawa, Ontario K1A 0R6, Canada

Hiroshi Sato, Cochair

National Inst. for Advanced Industrial Science Technology, Inst. for Human Science and Biomedical Engineering, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566 Japan

Contributed Papers

1:00

5pAA1. Calibration of consonant perception in room reverberation.
Kanako Ueno (Inst. of Industrial Sci., Univ. of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan), Norbert Kopco, and Barbara Shinn-Cunningham (Boston Univ., 677 Beacon St., Boston, MA 02215)

Many studies of sound perception often assumed that our auditory sensory processes are relatively static, rather than plastic. However, in everyday environments, we naturally and fluidly compensate for interfering effects of background noise and room reverberation. In order to investigate how listeners calibrate auditory perception to such acoustic interference, a listening experiment was performed to measure the effect of sudden changes of reverberation on the identification of consonants. Test sounds were generated by convolving two types of binaural room impulse responses (BRIRs) measured in large real rooms with speech tokens. As a control condition, pseudo-anechoic BRIRs with negligible reverberation energy were used. Listeners were asked to identify the consonant in a vowel-consonant target. The target was preceded by a carrier phrase consisting of vowel-consonant pairs from the same talker. In some cases, the target and carrier phrase were processed by the same BRIRs, while in others the BRIR's processing target and carrier differed. Consistent effect of calibration was observed in one of the simulated rooms, but not in the other, suggesting that the ability to compensate for the effects of reverberation depends on the specific pattern of reverberation produced in a given room. [Work supported by AFOSR and NSF.]

1:15

5pAA2. Sentence context influences vowel perception in reverberant conditions. Janine Wotton (Dept. of Psych., Gustavus Adolphus College, 800 W. College Ave., St. Peter, MN 56082, jwotton2@gac.edu), Kristin Welsh, Crystal Smith, Rachel Elvebak, Samantha Haseltine (Gustavus Adolphus College, St. Peter, MN 56082), and Barbara Shinn-Cunningham (Boston Univ., Boston, MA 02115)

Sentences recorded with a Mid-western accent were convolved with head-related impulse responses that included different room reverberation conditions. The stimuli were presented binaurally through headphones in an echo-attenuated chamber and subjects ($n=23$) typed the sentences they heard. The target word was one of a vowel pair (cattle/kettle, jam/gem, gas/guess, past/pest) embedded as the second word in one of three sentence types. The neutral sentence provided little context for the word. Target words in sentences that provided strong contextual cues could be congruent or incongruent with the expectations of the subject, for example, "The cattle/kettle grazed in the meadow." Subjects made significantly more errors in the incongruent sentences compared to the neutral (Wilcoxon$=3.572$ $p<0.05$) or congruent sentences (Wilcoxon$=3.56$ $p<0.05$). When the target word was in a congruent sentence subjects performed equally well in reverberant or pseudo-anechoic conditions (Wilcoxon$=1.298$) but they made more errors in the reverberant condi-

tion for both neutral (Wilcoxon$=3.359$, $p<0.05$) and incongruent sentences (Wilcoxon$=2.241$, $p<0.05$). Results suggest that reverberation may cause listeners to rely more heavily on linguistic context to determine word meaning. [Work supported by NOHR, AFOSR.]

1:30

5pAA3. Perceptual compensation for reverberation: Effects of noise-context bandwidth. Simon J. Makin, Anthony J. Watkins, and Andrew P. Raimond (School of Psych., The Univ. of Reading, Earley Gate, Reading RG6 6AL, UK, s.j.makin@reading.ac.uk)

Perceptual compensation for reverberation is observed when the reverberation is applied to a test word (from a "sir"-to-"stir" continuum) and its context (e.g., "next you'll get_to click on") are varied independently. Increasing reverberation in test words decreases listeners' "stir" responses, as reverberation "fills the gap" that cues the [t]. Compensation occurs when the context's reverberation is commensurately increased, and "stir" responses increase back to the level found with minimal test-word reverberation. Compensation is strongest with speech contexts but also occurs with some noise-like contexts, including "signal-correlated noise" that has the wideband temporal envelope of the original speech. Also effective is a wideband noise that is given the temporal envelope seen at the output of a single auditory filter in response to speech. A narrow-band version of this "auditory-filtered" noise is not effective, but when contexts are made by summing of three or five of these bands, their effectiveness increases correspondingly. Compensation appears to be informed by the "tails" that reverberation adds at offsets, so it merely requires contexts with suitable temporal-envelope fluctuations. However, effects seem confined to the context's frequency region, so the crucial offsets need to be in a wide range of frequency bands. [Work supported by EPSRC.]

1:45

5pAA4. Aural localization of speech stimuli. Evelyn Way (Talaske, 105 N. Oak Park Ave, Oak Park, IL 60301, evelyn@talaske.com)

Localization error was tested for a variety of signals to answer the question: do humans aurally localize different speech stimuli with a different localization blur? A series of tests was conducted comparing the effect of the sentence length, gender of the talker, and frequency content of the signals on localization. Results were applied to ongoing research into constructing an aurally accurate telepresence system.

3358   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3358**

2:00

5pAA5. Quantifying the effects of room acoustics on speech intelligibility in very low signal-to-noise ratio environments. Jarrod E. Whittington (Grad. Program in Archictectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180, whittj@rpi.edu) and John S. Bradley (Inst. for Res. in Construction, Ottawa, K1A OR6 Canada)

The intelligibility of speech transmitted from closed offices to adjacent spaces is strongly affected by the signal-to-noise ratio at the receiver position and the acoustical characteristics of the spaces involved. Previous studies have suggested that the effect of room acoustics on speech intelligibility in closed offices and rooms is negligible and can be ignored (as with intelligibility quantifiers such as the articulation index). The purpose of this study is to show that in conditions of very low signal-to-noise (i.e., when high speech privacy is a primary concern), the influence of room acoustics rises dramatically. To this end, multiple subjects were given tests of speech intelligibility in simulated sound fields. Speech samples were presented to subjects with seven levels of signal-to-noise and four different reverberation times. The results from these tests show that as reverberation time increases, speech intelligibility decreases much more sharply for very low (-8 dB) signal-to-noise situations than in higher ($+10$ dB) signal-to-noise situations. This suggests an important relationship between room acoustics and speech privacy/security.

2:15

5pAA6. The effect of a preprocessing approach improving speech intelligibility in reverberation considering a public-address system and room acoustics. Nao Hodoshima, Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., Tokyo 102-8554, Japan), and Peter Svensson (Norwegian Univ. of Sci. and Technol., Trondheim NO-7491, Norway)

This study evaluates a preprocessing approach for reducing reverberation effects when the input signal is not ideal, dry speech, but rather a realistic speech signal picked up by a close microphone in a room. And this study shows how the situation affects the input and the further approach compared to a dry signal as the input. Steady-state suppression, as described by Arai *et al.* [Acoust. Sci. Technol. **23**, 229–232 (2002)], was used as a preprocessing approach that processes a speech signal before it is radiated from loudspeakers. A lecture was simulated in two different halls (reverberation times of 1.2 and 1.8 s) in which public address systems were installed. The simulation software CATT-Acoustic was used and impulse responses were calculated for the input to the preprocessing approach and for a listener position. Stimuli for a syllable identification test were prepared by convolving speech signals with the calculated impulse responses. Speech signals were given with and without steady-state suppression. The inclusion of natural and electroacoustical impulse responses makes the study of steady-state suppression more realistic and tests its robustness. [Work supported by JSPS (176911).]

2:30

5pAA7. Improving speech intelligibility in reverberant rooms. Douglas F. Winker and Elmer L. Hixson (Dept. of Elec. and Comp. Eng., Univ. of Texas, Austin, TX 78712)

Speech intelligibility in rooms with long reverberation times has long been a problem for acousticians. This problem is exacerbated when the option for acoustic treatments is not an option for a variety of reasons. In this study, a constant-beamwidth, wide-bandwidth (CBWB) loudspeaker array was used to improve speech intelligibility in two multipurpose rooms where treatment was not an option. A CBWB loudspeaker array was designed with independent beam pattern control on both the horizontal and vertical axes. A wide horizontal beam pattern and narrow vertical pattern were designed to achieve constant coverage from 500 Hz to 4 kHz. The array incorporates four separate arrays, two with nested elements, and linear-phase FIR filters to maintain constant beamwidth over the three octaves of interest. Room simulations were conducted with CATT-Acoustic and the array was implemented and compared to a source with a

more conventional coverage pattern. The STI method was used to measure speech intelligibility at several positions throughout the rooms. The study showed an improvement in speech intelligibility and a reduction in measured reverberation times.

2:45

5pAA8. Experimental investigations of the influence of room acoustics on the teacher's voice. Malte Kob, Anja Kamprolf, Christiane Neuschaefer-Rube, Oliver Goldschmidt (Chair of Phoniatrics and Pedaudiology, RWTH Aachen Univ., Pauwelsstr. 30, 52074 Aachen, Germany, mkob@ukaachen.de), and Gottfried Behler (RWTH Aachen Univ., 52066 Aachen, Germany)

Teachers belong to the group of professional voice users who often suffer from voice disorders. One reason for a significantly increased prevalence of voice problems can be poor room acoustical conditions in the classrooms. In this study, four rather reverberant and loud classrooms in a primary school in Aachen were analyzed using measurements of the reverberation time, T_{20}, and the speech transmission index, STI. About half of the school's teachers were investigated with respect to their voice status by using phoniatric, logopedic, and objective voice analysis methods. The prevalence of voice problems in this group was found to exceed previous studies where subjective voice quality was rated. In a second part of this joint project the change of voice quality during the teachers' working day was analyzed. Two of the four rooms were acoustically optimized. Members of two groups of teachers with and without voice problems were recorded before and after teaching in either one of the acoustically poor rooms or one of the newly renovated rooms. The preliminary results indicate changes of the voice quality in most subjects with respect to one or more voice parameters. Further studies shall prove the significance of the room influence.

3:00–3:15 Break

3:15

5pAA9. Human voice phoneme directivity pattern measurements. Brian F. G. Katz, Fabien Prezat, and Christophe d'Alessandro (Percept. Située, LIMSI-CNRS, BP 133, F91403 Orsay, France)

The application of directivity patterns to radiating sources into computer simulations and auralizations is common for loudspeaker models. Few applications include the directivity patterns of natural sources, partly due to the lack of sufficient data. This work presents the results of a detailed measurement study on human voice directivity in three-dimensions. Unlike previous studies that have used average directivity data over read phrases, this work presents results that are measured for a number of sustained individual phonemes. Details of the measurement protocol and posttreatment processing are presented. Comparisons are made relative to phoneme, f0, spectral characteristics, and associated mouth geometry for several talker subjects. Studies have also been made on the directivity of the singing voice. Specifically, the variations in directivity relative to level (piano, fortissimo, etc.) and projection as controlled by the singer have been investigated. Results of this work are applicable to speech production research, talker simulator design, room acoustic sound field prediction, and virtual reality systems with talking avatars.

3:30

5pAA10. Case study for voice amplification in a highly absorptive conference room using a negative absorption tuning by the Yamaha Active Field Control system. Takayuki Watanabe (YAMAHA Commercial Audio Systems, Inc., 6600 Orangethorpe Ave., Buena Park, CA 90620)

The Yamaha Active Field Control (AFC) enhances and varies acoustical conditions from the room acoustics using the acoustical feedback in a system, which is properly tuned to the installed room acoustics. The system is primarily used to improve auditory impressions of reverberance, loudness by enhancement of reverberation in time, and energy in a music venue. Here, a negative absorption tuning is demonstrated as beneficial to

the presentation room for a simple spoken speech used with a relatively excessive absorption at a middle-sized multiuse conference room. The space occupies 1100 ft² with a ceiling height of 9 ft. It is designed for demonstration of sound equipment as well as lecture without audio equipment. Since the acoustics in the room are highly absorptive (average absorption coefficient, 0.3), considering a primary use for product demonstration, it causes a flutter echo between longitudinal walls due to less sound energy in a late field. We demonstrate here a system adjustment of the AFC system emphasizing voice amplification as well as an acoustical solution to a defect of a fluttering echo here. Acoustical descriptors such as RT, echo diagram, and SPL are discussed for a tuned room acoustics with the AFC system.

Contributed Poster Papers

Poster papers 5pAA11 to 5pAA14 will be on display from 1:00 p.m. to 4:45 p.m. Authors will be at their posters from 3:45 p.m. to 4:45 p.m.

5pAA11. Slowed speech spreading into reverberant environments; steady-state suppression improves speech intelligibility. Yuki Nakata, Yoshiaki Murakami, Nao Hodoshima, and Takayuki Arai (Dept. of Elec. and Electrons Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan)

Speech intelligibility is often decreased in reverberant environments. For improving speech intelligibility under reverberant environments, Arai *et al.* [Acoust. Sci. Technol. **23**, 229–232 (2002)] suggested steady-state suppression, which suppresses steady-state portions of speech to reduce overlap-masking (causing degradation intelligibility) and improves speech intelligibility as preprocessing. Previous studies showed that speech intelligibility was improved through this processing under some reverberant environments [Hodoshima *et al.*, J. Acoust. Soc. Am. **119**(6), 4055–4064 (2006)]. On the other hand, speech intelligibility is known to be increased greatly by speaking slowly [Bolt and MacDonald, J. Acoust. Soc. Am. **21**(6), 577–580 (1949)]. However, it is not optimal for improving speech intelligibility in terms of reducing overlap masking because vowels causing major parts of overlap masking are expanded by speaking slowly. Therefore, in the current study, we investigated the effects of steady-state suppression on speech perception with a decreased speech rate from 6 to 4 and 5 morae/s. Results showed that the slowest speech (4 morae/s) with steady-state suppression was the most intelligible. Also, steady-state suppression improved speech intelligibility at a speech rate of 4 morae/s in each reverberant condition (reverberation times of 1.5, 2.0, and 2.5 s). [Work partially supported by JSPS.KAKENHI (16203041).]

5pAA12. Suppression of speech intelligibility loss through a modulation transfer function-based speech dereverberation method. Masashi Unoki, Masato Toi, Yohei Shibano, and Masato Akagi (School of Information and Sci., JAIST, 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan, unoki@jaist.ac.jp)

The concept of modulation transfer function (MTF) can successfully be applied to evaluate the quality of speech transmission in the room acoustics [Houtgast and Steeneken, J. Acoust. Soc. Am. **77**, 1069–1077 (1985)]. We previously proposed a speech dereverberation method based on the MTF concept, which consisted of MTF-based power envelope inverse filtering and the carrier regeneration in the filterbank [Unoki *et al.*, EuroSpeech2003 (2003)]. This paper evaluates how the proposed method can suppress the loss of speech intelligibility caused by reverberation, by comparing various methods. We have carried out massive simulations of dereverberation for reverberant speech signals to objectively evaluate these methods. We also subjectively evaluated the methods via listening tests. In these simulations, artificial reverberations were convolved with a clean speech signal in which the impulse responses in the room acoustics can be approximated from the exponential decay as a function of reverberation time with a white-noise carrier. The results of both evaluations show that, in addition to reducing the averaged log-spectrum distortion by about 1 dB, the proposed method reduces the loss of speech intelligibility by about 30%. [Work supported by a Grant-in-Aid for Science Research from the Japanese Ministry of Education (No. 18680017).]

5pAA13. Word and mora intelligibility in "Familiarity-controlled word-lists 2003 (FW03)." Tadahisa Kondo, Shigeaki Amano (NTT Commun. Sci. Labs., NTT Corp., 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa, 2430198, Japan, tkondo@brl.ntt.co.jp), Shuichi Sakamoto, and Yôiti Suzuki (Tohoku Univ., Aoba-ku, Sendai, Miyagi, 9808577, Japan)

"Familiarity-controlled Word-lists (FW03)" was developed to make it possible to perform intelligibility tests on the same person repeatedly and/or under several different conditions. FW03 consists of 20 lists of 50 words in four word familiarity ranks. These words were selected taking phonetic balance into consideration to maximize their variety of initial moras and vowel-consonant sequences. To confirm that all FW03 word lists from a particular familiarity rank present the same hearing difficulty, we measured the word intelligibility of all the words in FW03. Sixteen subjects listened to 4000 words spoken by four speakers with seven signal-to-noise ratios. The mean intelligibility scores for the lists were significantly different even for the same familiarity rank. The intelligibility was apparently influenced by some specific moras and their position in words. These results suggest that it is difficult to equalize the word intelligibility of the lists even when the word familiarity is controlled more precisely. We therefore examined several methods for equalizing the intelligibility among the lists such as controlling the sound pressure level for each word when operating an actual intelligibility test. The effectiveness of these methods is discussed with detailed analyses of word and mora intelligibility in FW03.

5pAA14. Influence of Deutlichkeit value and reverberation time on improved speech intelligibility in reverberant environments because of steady-state suppression. Nahoko Hayashi, Nao Hodoshima, Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554 Japan, nhayash@sophia.ac.jp), and Kiyohiro Kurisu (TOA Corp., Takarazuka, Hyogo 665-0043 Japan.)

To improve speech intelligibility in reverberant environments, Arai *et al.* proposed "steady-state suppression (SSS)" as preprocessing [Arai *et al.*, Acoust. Sci. Technol. **23**, 229–232 (2002)]. In this study, a perceptual experiment under artificial reverberant conditions with simulated impulse responses was conducted to elucidate the effect of the Deutlichkeit (D) value and reverberation time (RT) on improvements of speech intelligibility because of SSS. Artificial impulse responses were simulated with white noise multiplied by a decay curve. The advantage of this method is that the simulated impulse responses have mutually similar frequency characteristics; consequently, we can evaluate them using only the D value and RT regardless of their different frequency characteristics. Two parameters, the energy of the impulse response 50 ms from the direct sound and the attenuation rate of the decay curve, were controlled to obtain several impulse responses having certain D value and RT. Results show that SSS improved speech intelligibility in the conditions of low D value, even if RT was long or short. We could also interpret these results as indicating that processing is effective when the original speech intelligibility is less than 60%. [Work supported by JSPS.KAKENHI (16203041).]

Session 5pBB

Biomedical Ultrasound/Bioresponse to Vibration: Medical Ultrasound

Yuri A. Pishalnikov, Cochair
Indiana Univ., Anatomy and Cell Biology, 635 Barnhill Dr., Indianapolis, IN 46202

Yoshiaki Watanabe, Cochair
Doshisha Univ., Faculty of Engineering, 1-3 Tatara-Miyakodami, Kyotamabe-shi, Kyoto 610-0321, Japan

Contributed Papers

1:00

5pBB1. Measurement of thrombi volume using a transesophageal echocoardiography movie. Masami Aoki (Grad. School of Sci. and Technol., Chiba Univ., 1-33 Yayoi-cho Inage-ku, Chiba, 263-8522, Japan), Tadashi Yamaguchi, and Hiroyuki Hachiya (Chiba Univ., Chiba, Japan)

In a total knee replacement operation, thrombi are formed by blood stasis in the deep venous of inferior limbs. They can cause pulmonary thromboembolism. The current judgment method of the existence of thrombi in the orthopedics field is visual observation using transesophageal echocoardiography (TEE). Therefore, we attempted to devise a quantitative measurement method of the volume of thrombi from TEE movies using multiple signal processing. It is necessary to extract a contour of the right atrium area at each frame of a TEE movie to measure the amount of thrombi that reach the lungs. Therefore, we propose processing that combines the modified brain extraction technique (BET) and vector moving method. The modified BET enables high-precision contour extraction. Vector moving performs the fast contour extraction. In addition, the fiber structure extraction technique (FSET) is used to extract the thrombi information in the right atrium. Similarly to speckle reduction, FSET can extract anomalous information from echo images. Using these processes, we observed a small periodic change that is synchronized with the heartbeat and the large time change that is synchronized with breathing.

1:15

5pBB2. Auscultation in high-noise environments using hybrid electromechanical and ultrasound-Doppler techniques. Adrianus J. M. Houtsma, Ian P. Curry (U.S. Army Aeromedical Res. Lab., Fort Rucker, AL 36362), John M. Sewell, and William N. Bernard (Active Signal Technologies, Inc., Linthicum Heights, MD 21090)

Medical evacuation of battlefield casualties or traffic accident victims by air typically takes place in very noisy environments. Auscultation of patients, e.g., to support chest intubation or to detect a pneumothorax, is therefore difficult or impossible to perform. A conventional acoustic stethoscope will not function very well in background noise levels beyond 80 dB. Electronic stethoscopes, in combination with mechanical impedance-matched transducer designs, can extend this range to about 90 dB, but this is not enough for helicopter noise levels that can reach 110 dB. The use of an ultrasound transmitter and receiver, however, provides an essentially noise-free auscultation channel since transportation vehicles do not produce acoustic energy at ultrasound carrier frequencies of 2–3 MHz. Clean and noise-free heart and breath sounds have been obtained in broadband noise fields of intensities as high as 120 dB. A hybrid stethoscope has been developed that allows auscultation by ultrasound-Doppler as well as electromechanical means. Pros and cons of making Doppler sounds subjectively similar to conventional sounds by nonlinear signal processing will be discussed, as well as potentially functional and meaningful aspects of Doppler signals that are not found in conventional stethoscope sounds.

1:30

5pBB3. Color and pulsed Doppler signatures of vascular bleeding. Wenbo Luo, Vesna Zderic, and Shahram Vaezy (Dept. of Bioengineering, Univ. of Washington, Seattle, WA 98195)

To develop ultrasound-based methods for detection and localization of internal vascular bleeding, we investigated the bleeding signatures using color and pulsed Doppler. Under ultrasound guidance, femoral arteries ($n = 13$) and veins ($n = 12$) in seven pigs were catheterized with 6, 9, and 11 F catheters. Color and pulsed Doppler were acquired from the injury site and the surrounding positions after withdrawal of the catheter. The bleeding rates were 3.8–38 ml/s. Color Doppler revealed that the injured vessels were in spasm. The checkered color pattern at the injury site indicated flow turbulence. Pulsed Doppler spectrum of the color jet (extravasated blood) represented a unique arterial bleeding pattern with increased baseline velocity (from 0 up to 25.8 cm/s) and elevated systolic velocity (from 23.2 up to 62.5 cm/s) at the puncture site, as compared to the normal arterial pattern. Venous bleeding showed a narrow-band vibration signal at the bleeding site. Local blood turbulence contained multiple velocity components which were illustrated as checkered color in color Doppler. Pulsed Doppler spectrum showed increased baseline due to turbulence and elevated velocity due to narrowed vessel. In future studies pulsed Doppler radio-frequency data will be used to analyze bleeding patterns for detection and localization of internal bleeding.

1:45

5pBB4. Basic study on angular dependence of ultrasonic scattering from wire phantom mimicking myocardium. Teppei Onodera, Hideyuki Hasegawa, and Hiroshi Kanai (Grad. School of Eng., Tohoku Univ., Sendai 980-8579, Japan, onodera@us.ecei.tohoku.ac.jp)

It is reported that the ultrasonic scattering from myocardium varies periodically during a cardiac cycle. One of the reasons is considered to be the change of the angle between the ultrasonic beam and the direction of myocardial fibers. Therefore, it is expected that ultrasonic scattering from myocardium quantitatively characterizes the condition of myocardium. To investigate the angle dependence of ultrasonic scattering in relation to the fiber direction, in this study, ultrasonic echoes from a wire phantom (thinner than a wavelength in diameter) that mimics myocardium were measured as a function of the insonification angle. Two ultrasonic transducers of 7 MHz were employed for transmitting and receiving ultrasonic pulses. Focal points of these transducers were set at the same point on the phantom. A custom-made experimental system can change the azimuth and elevation angles while keeping the focal points at the same point. Measured ultrasonic scattering showed a significant change depending on the azimuth angle. Such changes depending on the azimuth angle were decreased by increasing the elevation angle. These preliminary findings support the hypothesis that the change in the angle between the ultrasonic beam and myocardial fibers due to the heartbeat varies the ultrasonic scattering property.

3361   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3361**

5p SAT. PM

Session 5pMU

Musical Acoustics: Simulation and Measurement Techniques for Musical Acoustics II

Shigeru Yoshikawa, Cochair

Kyushu Univ., Dept. of Acoustic Design, 4-6-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

Gary P. Scavone, Cochair

McGill Univ., Music Technology, Schulich School of Music, 555 Sherbrooke St., W., Montreal, Quebec H3A 1E3, Canada

Murray D. Campbell, Cochair

Univ. of Edinburgh, School of Physics, Mayfield Rd., Edinburgh EH9 3JZ, U.K.

Invited Papers

1:00

5pMU1. A hybrid approach for simulating clarinetlike systems involving the lattice Boltzmann method and a finite difference scheme. Andrey Ricardo da Silva and Gary Scavone (Computational Acoust. Modeling Lab., McGill Univ., 555 Sherbrooke St. W., Montreal, CA H3A 1E3, Canada, andrey.dasilva@mail.mcgill.ca)

The lattice Boltzmann (LB) method is becoming an increasingly important technique for the simulation of problems involving the interaction between acoustic and fluid fields. New LB algorithms allow the representation of moving boundaries at low Mach numbers. In this work we present the implementation of a hybrid model involving one such LB algorithm coupled to a finite difference scheme to simulate the fluid-structure interaction at the reed of a two-dimensional clarinetlike system. The results of the simulation for different input pressures and reed parameters are compared with the results provided by the related literature. [The first author was supported by CAPES (Brazil).]

1:20

5pMU2. Three-dimensionality and temporal variability of the jet flow in organ pipes. Shigeru Yoshikawa, Yumiko Sakamoto (Dept. of Acoust. Design, Grad. School of Kyushu Univ., Fukuoka 815-8540 Japan), Andreas Bamberger (Albert-Ludwigs-Univ. Freiburg, D-79104 Freiburg, Germany), and Judit Angster (Fraunhofer Inst. for Bldg. Phys., D-70569 Stuttgart, Germany)

Flow visualization by the particle image velocimetry (PIV) is expected to provide quantitative information based on high resolution from the use of sheetlike laser pulses. This PIV technique is applied to investigate the difference in jet-flow structure of organ pipes when the ear (small plate standing at each side of the mouth) is used or not. The experimental results given by PIV are also compared with those by the hot-wire anemometry. Two-dimensionality and laminar-flow condition tend to be maintained when the ear is used. However, when the ear is not applied, the deviation from laminar flow and the resulting three-dimensionality are apparent halfway between the flue and the edge (*re* less than 1800). The ear seems to block the entrainment from the mouth side. Moreover, it is shown that the magnitude of jet velocity is significantly reduced at the phase when the jet deflects innermostly to the pipe with the ear and at the phase when the jet deflects outermostly from the pipe without the ear. These 3D and temporal variability of the jet flow should be examined in more detail in the context of harmonic structure of organ tones and vortex-sound theory on flue instruments.

1:40

5pMU3. Vortex sound and the flute. Andreas Bamberger (Phys. Institut der Universitaet Freieburg, Hermann Herder Str. 3, 79104 Freiburg, Germany)

The flute is investigated for its aeroacoustical properties based on the vortex sound theory. Particle image velocimetry (PIV) is used for quantitative flow determination of the jet-edge interaction. The Endoscopic-PIV offers a nonobstructive view of the system over all phases. The evaluation of the source term through the vorticity is done according to M. Howe (1975). The acoustic flow across the embouchure is determined by excitation through the foot. The flute is operated near 1200 Hz with various jet speeds. Finally the acoustic radiation power in the far field is determined to be compared with the source terms. The following findings are presented: (*i*) The space integrated and time averaged power of the coherent source terms turns out to be positive, i.e., emitting acoustic energy. (*ii*) There is a dominant contribution near the labium. (*iii*) The source term power compares favorably with the far field power, which is well below 1% of the total input power. The results demonstrate that the vorticity of the upper and lower shear layers of the jet cancels to a great extent, except for regions where the transverse amplitude of the jet is large.

2:00

5pMU4. Sound generation in flutelike instruments: Simulation and analysis. Helmut Kuehnelt (Inst. of Musical Acoust., Univ. of Music and Performing Arts Vienna, Singerstrasse 26a, A-1010 Vienna, Austria)

Three-dimensional flow-acoustic simulations using the lattice Boltzmann method were used to study the sound generation in flutes and recorders. Theory suggests that the fluctuating Coriolis force in the mouth of the instrument caused by the fluctuating vorticity distribution of the jet is the main source of sound. A numerical experiment was undertaken to clarify this hypothesis for flute instruments. In the mouth of the instrument an equivalent temporal fluctuating force was imposed in a quiescent fluid. Although this force excited fluid dynamic effects in the mouth not present at mean flow, it nevertheless induced a standing wave in the instrument with a sound pressure whose amplitude differed only less than 1 dB from the playing situation. This shows the cardinal role of the Coriolis force for the sound generation in flutes. The temporal and spatial structure of the Coriolis force distribution and the vortex sound power generated will be compared for flute and recorder. Especially the spatial structure differs systematically for the two types of instruments. Implications for the design of flute head joints will be discussed.

Contributed Paper

2:20

5pMU5. Particle image velocimetry measurement and vortex sound analysis of high-amplitude acoustically generated flows in open ended tubes. David Skulina (Acoust. and Fluid Dynam. Group, School of Phys., Univ. of Edinburgh, Scotland)

Periodic vortex shedding induced by an acoustic field has been observed by several authors at orifices and at the open ends of tubes. It is a well-known dissipative mechanism and affects the sound field at the open end of a tube and the efficiency with which a standing wave can be maintained within it. This has obvious implications in the playability and design of musical wind instruments. To quantify the losses associated with vortex generation and shedding at the open end of a tube a method for the evaluation of PIV velocity maps using vortex sound theory is outlined. Comparison between the results of the vortex sound calculations and the velocity fields measured from the PIV vector maps allows both qualitative and quantitative descriptions of the acoustical effect of vortex shedding on losses over the acoustic cycle. It is found that the net acoustical losses due to the interaction of vortices with the acoustic field are most significant in a region within 2.5 mm of the open end. Out of this region, vortices act as a net source of sound, but are of lesser magnitude. [Work supported by EPSRC.]

2:35–2:50   Break

Invited Papers

2:50

5pMU6. Synthetic percussion. Stefan Bilbao (Dept. of Music, Rm. 7306B, James Clerk Maxwell Bldg., King's Bldg., Mayfield Rd., Edinburgh EH3 9JZ, Scotland)

The widespread availability of fast small computers has removed many of the restrictions that governed the development of sound synthesis methods in the past, and, in particular, those based on physical modeling of musical instruments. Even relatively complex models can now produce sound output in something approaching real time, if not real time itself. As an example, various models of strings and 2-D percussion instruments, simulated using direct time domain methods, are presented here. Various issues are discussed, including interconnections of various vibrating objects, extensions of percussion instruments involving acoustic "preparation," the modeling of nonlinearities, multichannel output, algorithmic complexity, and operation counts, as well as current work towards developing real time implementations. Sound examples are presented.

3:10

5pMU7. Simulation of piano strings and a soundboard by a large deformation theory. Isoharu Nishiguchi, Masataka Sasaki, and Aki Yamamoto (Kanagawa Inst. of Technol., 1030 Shimoogino, Atsugi, Kanagawa, Japan 243-0292, nishiguc@sd.kanagawa-it.ac.jp)

Three-dimensional finite element analyses of piano strings and a soundboard are presented in which a large deformation theory is adopted. In the large deformation theory, the momentum balance of a body is preserved in the course of the deformation and the changes of direction and amplitude of axial force due to deformation are taken into account. It is shown that the secondary partials appear in the spectrum of the velocity component in the axial direction as well as the transverse direction. A hammer is modeled by masses with springs and dashpots. Gap elements are employed to simulate the contact and detachment between the hammer model and a string. It is demonstrated that multiple reflections between hammer and agraffe can be reproduced by the method. Modeling of the interaction among strings, a bridge, and a soundboard is also discussed in which the crown and the preloading of the soundboard are taken into account.

3:30

5pMU8. Physical modeling of the piano. N. Giordano (Dept. of Phys., Purdue Univ., West Lafayette, IN 47907, giordano@purdue.edu)

As part of an ongoing project we have constructed a computational model of the piano. This model uses Newton's laws to compute the motion of the hammers, strings, soundboard, and room air, yielding the sound pressure that reaches a listener. The model is based heavily on corresponding experimental studies of hammers, strings, and soundboard properties. To date we have focused on the modern piano, but our basic modeling approach can also be used to investigate how changes in the instrument would change the resulting tones. After reviewing the model, we describe studies of the sound produced by models of historical pianos, i.e., the type of piano made in the late 1700s or early 1800s, and compare those tones with the tone of a modern instrument.

5pMU9. Numerical dispersion of finite difference method. Yoshimitsu Takasawa (Univ. of Electro-Commun. 1-5-1 Chofugaoka, Chofu, Tokyo, Japan, takasawa@ice.uec.ac.jp)

The partial frequencies of the transverse vibrations of a string, which has stiffness ϵ and is assumed to be hinged at both ends, form an inharmonic series and obey the equation $f_n = f_0 n (1 + \pi^2 \epsilon n^2)^{1/2}$, where f_0 is the fundamental frequency. For musical synthesis, the equations of a physical model can be simulated in the time domain using numerical methods. The most straightforward approach can be said to be the finite difference method. The equations are formulated in discrete form using discrete positions $i\Delta x$ and discrete time steps $n\Delta t$. In the finite difference method, two problems must be taken into account. The first is numerical stability: it must be roughly $\gamma > N$. Here, $\gamma = f_e/2f_0$, $f_e = 1/\Delta t$ is the sampling frequency, and $N = L/\Delta x$ is the segment number of the string. The second is the numerical dispersion. It has been shown that some undesirable dispersive effects might be present in the solution if a finite difference scheme is used. This numerical dispersion can be formularized as the equation $f_n \approx f_0 n [1 + \pi^2 \{\epsilon - \frac{1}{12}(1/N^2) - 1/\gamma^2)\} n^2 - \pi^4/6\{\epsilon/N^2 - \frac{1}{60}(1/N^4 - 1/\gamma^4)\} n^4]^{1/2}, n \leq N-1$.

5pMU10. Imaging vibrations and flow using electronic speckle pattern interferometry. Thomas R. Moore (Dept. of Phys., Rollins College, Winter Park, FL 32789)

Electronic speckle pattern interferometry has been used to measure sub-micrometer displacements and visualize the deflection shapes of harmonically vibrating objects for many years; however, to date the technique has not been widely used within the community of musical acousticians. The cost of the equipment, the need for effective vibration isolation, and the complexity of the hardware are usually cited as the reasons why this technique has not been more widely utilized in the study of musical instruments. To alleviate these concerns, a speckle pattern interferometer has been designed that is relatively inexpensive, easy to assemble, and very robust. Construction of the interferometer will be discussed, and applications to problems in musical acoustics will be presented.

5pMU11. Psychoacoustic experiments with virtual violins. Claudia Fritz, Ian Cross (Faculty of Music, Univ. of Cambridge, West Rd., Cambridge, CB3 9DP, UK), Brian C. J. Moore, and Jim Woodhouse (Univ. of Cambridge, Cambridge, CB3 9DP, UK)

This study is the first step in the psychoacoustic exploration of why some violins sound better than others. A method was used that enabled the same performance to be replayed on different "virtual violins," so that the relationships between acoustical characteristics of violins and perceived qualities could be explored. Recordings of real performances were made using a bridge-mounted force transducer, giving an accurate representation of the signal from the violin string. These were then played through filters corresponding to the admittance curves of different violins. Initially, the limits of performance in detecting changes in acoustical characteristics have been characterized. Thresholds were measured for the detection of different modifications of a violins acoustical response, such as a shift in frequency or an increase in amplitude of one or several modes, using a three-interval forced-choice discrimination task. Thresholds were higher for an input of a musical phrase than for a single note, but depended strongly on the choice of the note. The lowest threshold corresponded to a simultaneous shift in frequency of 1.5%, or an increase in level of 3 dB, of several modes, but thresholds appear to be dependent on the musical training of the listeners.

5pMU12. Use of the energy decay relief (EDR) to estimate partial-overtone decay times in a freely vibrating string. Nelson Lee and Julius O. SmithIII (Ctr. for Comput. Res. in Music and Acoust. (CCRMA), The Knoll, 660 Lomita, Stanford, CA 94305-8180, USA, nalee@stanford.edu)

The energy decay relief (EDR) was proposed by J. M. Jot [IEEE, ICASSP (1992)] for displaying the impulse response of artificial reverberation systems. The EDR is a frequency-dependent generalization of Manfred Schroeder's energy decay curve (EDC), defined at time n as the sum of squared impulse-response samples from time n until decay is complete. The EDR is similarly defined for each band in a uniform filter bank, typically implemented using the short-time Fourier transform (STFT). In this work, we apply the EDR to the problem of estimating decay times for the partials of a freely vibrating string. Previously, such decay times have been measured based on STFT magnitude data. We show that the EDR has certain advantages over the STFT, such as being less sensitive to "beating" in the amplitude envelopes of the partial overtones. Results in the context of virtual acoustic guitar modeling will be presented.

Contributed Paper

5pMU13. Numerical analysis of the 48-key experimental piano. Kenji Hasegawa, Isoharu Nishiguchi, and Masataka Sasaki (Kanagawa Instutute of Technol., 1030 Shimoogino, Atsugi, Kanagawa, 243-0292, Japan, s064817@cce.kanagawa-it.ac.jp)

An experimental piano has been built as a part of research program for students in the 4th grade. Though the main purpose of the building was to use the piano in vibrational and acoustical experiments, it was also intended to deepen the understanding of the structures of a piano by newly designing and building a piano that was different from existing pianos. The range covered by the piano is 4 oct (key nos. 21–68). After the sizes and arrangements of the strings were determined, the metal frame that sustains the tension of the strings was designed. The deformation and the stress were evaluated by the FEM. In this paper, the design and results of numerical analysis of the strings and soundboard are reported.

Session 5pNSa

Noise and Physical Acoustics: Prediction and Propagation of Outdoor Noise II

D. Keith Wilson, Cochair

U.S. Army Cold Regions Research Lab., Engineering Research and Development Ctr., 72 Lyme Rd., Hanover, NH 03755-1290

Kohei Yamamoto, Cochair

Kobayashi Inst. of Physical Research, 3-20-41 Higoshi-Motomachi, Kokubunnji-shi, Tokyo 185-0022, Japan

Contributed Papers

1:00

5pNSa1. A predictive noise model for the resurfacing of Nebraska Highway 2. Dominique Cheenne, Connie Lee, Michael Cappiello, Sam Lalk, Coleman Martin, and Philip Muzzy (Dept. of Audio Arts & Acoust., Columbia College Chicago, Chicago, IL 60610)

A detailed investigation of the noise soundscape resulting from traffic on a 3-mile segment of Nebraska Highway 2 prior to resurfacing work was undertaken by the students enrolled in the Acoustics Program at Columbia College, Chicago. Noise data were acquired at six different locations using established protocols and included daytime and nighttime conditions. The test data were compared to the predictions from an outdoor noise propagation computer model and the results were found to be within ± 2 dB for all locations under consideration. The computer model was then used to assess expected noise conditions under various concrete and asphalt scenarios for resurfacing. Data show that reductions of up to 6 dB in the overall noise levels may be achievable depending on the choice for the surface of the roadway.

1:15

5pNSa2. Experimental determination of the correction for noise levels assessed in front of a facade. Gianluca Memoli (Memolix Environ. Consultants, Pisa, Italy), Stylianos Kephalopoulos, Marco Paviotti (Joint Res. Ctr. of the European Commission, Ispra, Italy), and Gaetano Licitra (Tuscany Region Agency for Environ. Protection, Firenze, Italy)

The assessment of noise levels, in the proximity of a building or on its facade, is a requirement of the European Environmental Noise Directive 2002/49/EC concerning environmental noise produced by road and railway traffic, airports, and industries. One of the problems that concerns both theoretical models and measurements is to estimate the correction due to reflections. Several works in the past have approached this problem, now stated by international standards, which suggest using a $+3$ or $+6$-dB correction in function of the microphone position. A complete set of measurements have been performed in the framework of the IMAGINE project for road traffic, in different urban situations, in order to determine

the real intereference pattern in terms of coherence length. A careful treatment of the acquired data highlighted a dependence of corrective factors on the distance from the building facade that cannot be neglected. Practical solutions and uncertainties also will be discussed in this work.

1:30

5pNSa3. An exact point-source starting field for the Green's function parabolic equation in outdoor sound propagation. Kenneth Gilbert and Xiao Di (Univ. of Mississippi, Natl. Ctr. for Physical Acoust., Coliseum Dr., Universtiy, MS 38677)

A method for exactly representing a point source in a Green's function parabolic equation calculation is presented. The method is no more difficult to use than standard starting-field formulations and has the several advantages: (1) the point-source solution is essentially exact; (2) only the propagating Fourier components are used; and (3) there is no limitation on the source height, which can be taken to be at $z = 0$, if necessary. The method is derived analytically and examples of the numerical implementation are given. [Research supported by U.S. Army Armament, Research Development and Engineering Center (ARDEC).]

1:45

5pNSa4. Predicted effects of forest stand age on acoustic propagation. Michelle Swearingen (Construction Eng. Res. Lab., P.O. Box 9005, Champaign, IL 61826, michelle.e.swearingen@erdc.usace.army.mil)

A theoretical study on the impact of forest stand age on acoustic propagation will be presented. A Red Pine forest was simulated at 10, 20, 30, 40, and 80 years of age, with assumptions that it was being maintained for utility pole harvest. Forest parameters, such as density, height, and diameter of trees, were used to predict vertical sound-speed profiles and then acoustic propagation. The resulting spectra were weighted to simulate low-frequency (30 Hz) and midfrequency (500 Hz) blast signatures. Spectra and sound-exposure levels (SEL) were examined to determine whether the forest stand age has a significant impact on acoustic propagation within a forest.

3365   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3365**

Session 5pNSb

Noise and ASA Committee on Standards: Consumer Product Noise

Matthew A. Nobile, Cochair

IBM Hudson Valley Acoustical Lab., M/S P226, Bldg. 704, Boardman Road Site, 2455 South Rd., Poughkeepsie, NY 12601

Takeshi Toi, Cochair

*Chuo Univ., Dept. of Precision Mechanics, Faculty of Science and Engineering, 1-13-27 Kasuga, Bunkyo-ku,
Tokyo 112-8551, Japan*

Chair's Introduction—2:15

Invited Papers

2:20

5pNSb1. Noise emissions from powered hand tools: A consumer alert!.　Charles Hayden and Edward Zechmann　(NIOSH, 4676 Columbia Pkwy., C27, Cincinnati, OH 45226)

　　Existing standards for acquiring sound power levels of powered hand tools specify, for the most part, that sound levels be gathered in the unloaded test condition. However, there is an increase in sound level when the power tool is tested in the loaded condition. There are two purposes for gathering sound power level data: (1) determination of hearing loss hazard or irritation and (2) part of an engineering design optimization effort. The former is the focus of the National Institute for Occupational Safety and Health efforts. Sound power level data were gathered in the loaded and unloaded conditions from a variety of electrically powered hand tools. A model was then used to estimate the A-weighted sound pressure level at the operator's ear from the A-weighted sound power (LWA) measurements of powered hand tools. The magnitude of LWA is found to be a reasonable predictor of the magnitude of sound pressure level exposure that a powered hand tool operator might experience across a variety of acoustical environments. This presentation explores the methods used to gather sound power levels, an evaluation of the model, and an examination of differences found between loaded and unloaded tool test conditions.

2:40

5pNSb2. Protocol for measuring dishwasher sound power levels.　Kevin Herreman and Richard Godfrey　(Owens Corning Corp., 2790 Columbus Rd., Rte. #16, Granville, OH 43023, kevin.herreman@owenscorning.com)

　　A leading retailer of home appliances came to the OC acoustics laboratory to evaluate the noise generated by dishwashers sold in their stores. A protocol based on a methodology used in Europe was used. It was modified to account for North American building standards. The standard provides an overall sound power level determination for the complete dishwasher cycle with and without the drying portion of the cycle. As a result customers making a purchasing decision can compare the noise performance of the various products using the same yard stick.

3:00

5pNSb3. Noise exposure limits and measurements for earphones.　K. Allen Woo　(Plantronics, Inc., 345 Encinal St., Santa Cruz, CA 95060)

　　There are noise exposure concerns regarding the use of audio earphones and telephone receivers. The permissible noise exposure limits have been defined for the occupational environment for many years, for instance, the US Code of Federal Regulations, 29CFR1910.95, Occupational noise exposure and the Directive 2003/10/EC of European Parliament on Minimum Health and Safety Requirements. Those permissible noise exposure limits are defined for noise exposure in an ''open-field'' environment. An open-field environment is an environment where the noise sources are at a distance from a person's ear. The sound or noise field can be a combination or any of a free field, or partially reflected, diffused, and reverberant fields. Nonetheless, the noise exposures from earphones are not in an open field. They are localized at or inside the user's ear. Conventional open-field noise exposure measurement methods do not apply. During the last 15 years or so, a few different test methods for measuring noise exposure from earphone or telephone receivers were developed. The standards of ITU-T P.360, UL 60950-1, IEEE 269, and EN 50332 were developed mostly by the telephone industry. This presentation addresses the common permissible noise exposure limits, how to relate the limits to earphones, and how to measure them.

3:20

5pNSb4. Sound quality evaluation of transient sounds in precision information equipments. Masao Yamaguchi (Res. & Development Ctr., Toshiba Tec Corp., 6-78, Minami-cho, Mishima-shi, Shizuoka-ken, 411-8520 Japan), Sho Shirakata, and Takeshi Toi (Chuo Univ., Bunkyo-ku, Tokyo 112-8511, Japan)

These days, sound quality in mechanical products has become a very important factor in addition to reduction of the sound pressure level. A new method is proposed for sound quality evaluation including the transient sound in precision information equipment taking MFP (multifunction peripheral), for example. The MFP has many functional parts and a lot of operational modes, so single copy mode and ten times continuous copies mode in monochrome are taken into account. Since the operational sound including the transient sound is changed with time in one cycle, it is necessary to identify high contribution sound section to decide the sound quality by pair comparison method and SD method with original, filtered, and edited sounds. As a result, sound quality of the transient sound is influenced by the cycle time, the rhythm, the sound pressure, and the frequency characteristics. As the single copy mode, main uncomfortable sound can be estimated at the steady-state sound section caused by a rotational component. On the other hand, for the ten times continuous copies mode, it can be estimated at the particular transient sound section in one cycle caused by collision with mechanical parts, which is different from the single copy mode.

3:40

5pNSb5. Sound source identification for precision information equipment having many kinds of sound sources. Sho Shirakata (Dept. of Precision Mech., Chuo Univ., 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan), Masao Yamaguchi (Toshiba Tec Corp., Shizuoka-ken, 411-8520, Japan), and Takeshi Toi (Chuo Univ., Tokyo 112-8551, Japan)

These days, precision information equipment having multifunction with high qualities in an office environment is demanded. Sound quality in this equipment has become a very important factor in addition to reduction of the sound-pressure level. MFP (multifunction peripheral) has many functional parts and operational modes, so it makes uncomfortable transient sounds in one cycle. The primary sound section to decide the sound quality of the MFP was estimated by sound quality evaluation as in previous research. Main uncomfortable sounds could be confirmed at the steady-state sound section as the single copy mode and at the particular transient sound section in one cycle as the ten-times continuous copies mode. So, it is necessary to identify the position of uncomfortable sound sources at each mode for comfortable sound design. Principal component analysis (PCA) using acceleration measured by accelerometers and sound pressure measured by microphones is applied to identify the number of sound sources. Sound-pressure distribution in one cycle from the MFP having many kinds of sound sources can be measured by microphone array at each moment. As a result, the sound source position can be identified by using PCA and sound-pressure distribution.

4:00

5pNSb6. The role of noise audits in characterizing and reducing consumer product noise. David L. Bowen (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

Consumer products often share some common traits that impact both the measurement techniques that can be employed, as well as the noise control options that are available. These traits include their relatively small size (except for the special case of automobiles), low profit margin, and quick time to market. Their small size implies that existing units or prototypes can usually be set up for acoustical measurements inside a modestly sized reverberation room. The low profit margin implies that design solutions, rather than adding ''acoustical'' materials, are the preferred route to a quieter product, but the quick time to market can sometimes work against this approach. This talk will focus on a sound power based ''noise audit'' technique that is useful for characterizing and rank ordering the potential noise sources in a product, in order to devise the most effective noise reduction strategies. Case studies will be used to illustrate the use of this engineering technique applied to the particular case of consumer products, where the design trend is often towards lighter, stiffer components and faster rotating parts, all of which tend to increase noise.

4:20

5pNSb7. Consumers, products, and noise: The economic, social, and political barriers to reducing noise in consumer products sold in North America. Leslie Blomberg (Noise Pollution Clearinghouse, Box 1137, Montpelier, VT 05601)

The paper examines the economic, social, and political barriers to reducing noise in consumer products sold in North America. Included in the economic issues are the cost of noise control and whether or not consumers are willing to pay more for a quieter product. An equally important economic barrier, however, is that the marketplace currently does not have an effective mechanism to reward builders of quieter products, even if noise treatments were cost effective. There is very little information available to consumers about the noise levels of consumer products. Without their informed decisions in the marketplace, no signal reaches the manufacturers. Complicating this problem is the consolidation of the consumer product manufacturing sector and the consolidation in the retail sector. The social barriers include the public's attitude towards noise (that it is an unavoidable cost of living in a technologically advance society) and the failure to see noise as a pollutant (that it is merely something unwanted). Political barriers include an absence of government involvement in consumer product noise, in terms of regulation, labeling, as well as a general concern for the issue.

4:40

5pNSb8. Evaluation of noise intensity level using an eccentric press during the production of metallic closure: Findings according to different positions surrounding the press. Paulo H. T. Zannin (Lab. de Acustica Ambiental, Industrial e Conforto Acustico, Dept. Eng. Mecanica, Univ. Federal do Parana, C. Politecnico, Curitiba, Pr, Brazil, 81531-990) and Samuel S. Ansay (Univ. Federal Tecnologica do Parana, Centro, Curitiba, Pr, Brazil, 80230-901)

This study was performed to evaluate the differences in noise level produced by an eccentric press, during the manufacturing of metallic clo-sures for steel cans. The primary objective of the study was to establish a rank of sound-intensity level and sound-power level as well as the global noise level emitted by this particular press. Ten different areas in the surface surrounding the press were evaluated regarding both the sound power and the sound intensity of each measurement. Each site was mea-sured twice to increase accuracy. The ten areas sampled were sweeping using intensity probe. We have found the surface area to demonstrate the lowest noise level was at S1 and the highest one was at S10 (p value 0.0005). In conclusion, our study demonstrates adequately the global noise level as well as in specific sites, indicating areas requiring special treat-ment for noise reduction in this environment.

SATURDAY AFTERNOON, 2 DECEMBER 2006

WAIANAE ROOM, 1:15 TO 5:15 P.M.

Session 5pPA

Physical Acoustics: Land Mines and Ultrasonics

Masahiro Ohno, Cochair

Chiba Inst. of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

Roger M. Waxler, Cochair

Univ. of Mississippi, National Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Contributed Papers

1:15

5pPA1. Optimizing seismo-acoustic sonar source array configuration for maximum radiated Rayleigh wave energy. Anthony N. Johnson (Dept. of Appl. Mathematics, Naval Postgraduate School, Monterey, CA 93943), Clyde L. Scandrett, and Steven R. Baker (Naval Postgraduate School, Monterey, CA 93943)

A three-dimensional (3-D) continuum mechanics approach to the de-velopment of a time-dependent finite element model for optimizing the position and excitation of a seismo-acoustic sonar source array to detect the presence of buried landmines will be presented. Various source con-figurations will demonstrate the use of constructive and destructive inter-ference, which maximizes the radiated energy of unidirectional Rayleigh waves while suppressing the radiation of body waves. Radiation charac-teristics are analyzed in a linear, horizontally stratified (isotropic and ho-mogeneous within each layer) half-space with a discrete number of tran-sient seismic sources. Results for Rayleigh wave strengths are presented in both a homogeneous half-space and a layered medium.

1:30

5pPA2. Nonlinear acoustic experiments involving landmine detection: Status and future. Murray S. Korman (Phys. Dept., U.S. Naval Acad., Annapolis, MD 21402) and James M. Sabatier (Univ. of Mississippi, University, MS 38677)

In past experiments [J. Acoust. Soc. Am. **116**, 3354–3369 (2004)], airborne sound at two primary tones, f_1 and f_2 (closely spaced near reso-nance), undergo acoustic-to-seismic coupling. Due to interactions with the buried plastic landmine, a scattered surface profile can be measured. Par-ticle velocity profiles at f_1, f_2, and $f_1-(f_2-f_1)$ exhibit single peaks; those at $2f_1-(f_2-f_1)$, f_1+f_2, and $2f_2+(f_2-f_1)$ involve higher-order mode shapes. For some combination frequencies the "on target" to "off target" contrast ratio can improve by over 20 dB. Near resonance, the bending (softening) of a family of increasing amplitude tuning curves, involving the vibration over the landmine, exhibits a linear relationship between the peak particle velocity and corresponding frequency. Hyster-esis effects and slow dynamic behavior have also been observed. The interaction between the top-plate interface of a buried landmine and the soil above it appears to exhibit many characteristics of the mesoscopic/nanoscale nonlinear effects that are observed in geomaterials like rocks (sandstone). In linear detection schemes the natural inhomogeneous soil layering can generate some false alarms. Recent "off target" versus. "on target" tuning curve measurements show a substantial increase in nonlin-earity, suggesting this to be a good false alarm discriminator. [Work sup-ported by ONR.]

1:45

5pPA3. Phase inversion and two-frequency interaction in nonlinear time-reversal acoustic method of land mine detection. Alexander Sutin (Artann Labs., Trenton / Stevens Inst. of Technol., Hoboken, NJ), Brad Libbey (U.S. Army RDECOM CERDEC NVESD, Fort Belvoir, VA 22060), and Armen Sarvazyan (Artann Labs., Inc., West Trenton, NJ 08618)

Seismic waves are known to generate a nonlinear response in soil, but data show that the presence of a mine increases the nonlinearities due to interactions of the mine and soil system. The nonlinear effects were inves-tigated by two methods. The first, phase inversion, uses two short, broad-band TRA focused signals with opposite signs. These are broadcast se-quentially and the responses are added in post processing. This results in cancellation of the linear response leaving the nonlinear system compo-nents. The second method relies on two-frequency intermodulation. In this case time-reversed seismic energy at 280 and 350 Hz was used to excite the system while the response was analyzed at the sum frequency, 630 Hz. The experiments demonstrate that the presence of the mine increases non-linearity by 12 dB for the phase-inversion method and by 6 dB for two-frequency interactions in comparison with intact sand. These results sup-port the hypothesis that the interface between mine body and surrounding sand significantly increases the nonlinear response as compared to sand alone. Work was supported by U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate.

5pPA4. Time-reversal acoustic focusing of waves produced by seismic vibrators. Alexander Sutin, Victor Kurtenoks (Artann Labs., Trenton/ Stevens Inst. of Technol., Hoboken, NJ), James Sabatier, Rick Burgett, Mike ONeil, Slava Aranchuk (Univ. of MS, University, MS 38677), Brad Libbey (U.S. Army RDECOM CERDEC Night Vision and Electron. Sensors Directorate, Fort Belvoir, VA), and Armen Sarvazyan (ARTANN Labs., Inc., NJ 08618)

A land mine detection system has been developed based on excitation of seismic waves by four mechanical shakers and measurements of ground vibration by a 16-beam, scanning laser Doppler vibrometer (LDV). Time-reversal acoustic (TRA) focusing was applied to concentrate seismic waves using the above system. Linear swept frequency signals and orthogonal initial signals in the frequency band of 100–500 Hz were used in outdoor testing. The system impulse response from any shaker to the LDV output was computed by cross correlating the original and recorded signals. With orthogonal signals, the system impulse responses between multiple shakers and the LDV were measured simultaneously, thus reducing the overall measurement time. Each shaker then reradiates the time-reversed impulse responses focusing the seismic energy to a point on the ground. Experiments demonstrate that the TRA focusing provided a high concentration of elastic wave energy in the test area with typical focusing spots about 20 cm in diameter. [Work supported by U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate.]

5pPA5. Acousto-optic interactions at a surface perturbed by acoustic and gravity waves. James Churnside, Hector Bravo (ESRL/Zeltech, 325 Broadway, Boulder, CO), Konstantin Naugolnykh, and Joseph Fuks (Univ. of Colorado, Boulder, CO)

We studied the acousto-optic interaction at a water surface perturbed by an underwater sound source. Using a coherent optical detection technique, we measured the power spectra of the scattered light. Filtering of the power spectra allowed us to detect the sound-induced perturbations of the surface in the presence of surface waves, even though the amplitude of the waves was much larger than the acoustic perturbations of the surface. The theory of scattering by a surface perturbed by sound and by surface waves is developed and that theory is compared with the experimental results.

5pPA6. Improvement of dispersion of nanometer-sized diamond particles for precise polishing by ultrasound exposure. Takeyoshi Uchida, Tsuneo Kikuchi (Natl. Metrology Inst. of Japan, Natl. Inst. of Adv. Industrial Sci. and Technol., Japan), Yoshikazu Setoyama, Norimichi Kawashima, and Shinichi Takeuchi (Toin Univ. of Yokohama)

A novel acoustic dispersion method for fine diamond particles was investigated by our group. Nanometer-sized diamond particles for abrasive agents should be dispersed sufficiently in water without aggregation. The particles usually aggregate immediately after their manufacture. The dispersants and dispersion equipment are used for dispersion of the particles. However, these dispersion methods present problems in their dispersion characteristics and reaggregation because such methods do not improve the essential surface characteristics. We suggested the acoustic dispersion method. We reported in 2005 at the IEEE Conference in Rotterdam that nanometer-sized diamond particles with 5-nm primary particle size were dispersed using ultrasound. The particles with 5-nm primary particle size are anticipated for use as future abrasive agents. Particles with 50-nm primary particle size are actually used at present as abrasive agents. Therefore, in this study, we investigated the dispersion of particles with primary particle sizes of 20, 50, 80, and 150 nm. Results show that the aggregated particles were disaggregated to nanometer-sized particles. Zeta potentials of the diamond particles were increased by ultrasound. These results suggest the improvement of dispersion of nanometer-sized diamond particles in water.

5pPA7. Observation of living samples in a C-mode configuration with an ultrasonic phase conjugator. Tatsuya Seki and Masahiro Ohno (Chiba Inst. of Technol., 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan)

Ultrasonic images of living samples have been obtained in a C-mode configuration with phase conjugation (time-reversal) processes. A phase conjugator for 10-MHz ultrasound was realized by the field-mixing in a PZT block pumped with 20-MHz electric fields. Samples, relatively thick (mm to cm) living materials, were placed between an ultrasonic transmitting/receiving transducer and the phase conjugator and were scanned to form C-mode transmission images. Ultrasonic tone-burst pulses radiated onto the sample were led into the PZT block and were converted to their phase conjugate waves, which, for their time-reversing nature, acted to cancel the wavefront deformation introduced by the thick living samples. In the experiment, images were obtained at a frequency of 10 MHz for several samples such as chicken liver, poke meat, and some model samples made of agarose gel, both by the phase conjugate method and by the conventional one. Images by the phase conjugate method showed relatively good accordance with the real ultrasonic attenuation distribution, while those by the conventional method were very sensitive to the shapes or the discontinuities of samples.

5pPA8. Propagation and interaction of phase conjugate ultrasound waves in moving media. V. Preobrazhensky, P. Pernod, and Yu Pyl'nov (Joint European Lab. LEMAC: Institut d'Electronique et de Micro-electronique et de Nanotechnologie IEMN-DOAE-UMR CNRS 8520, Ecole Centrale de Lille, BP48, 59651 Villeneuve d'Acsq, Cedex, France, Wave Res. Ctr. of General Phys. Inst. RAS and Moscow State Inst. of Radio Eng., Electron. Automation, Moscow, Russia, vladimir.preobrajenski@iemn.univ-lille1.fr)

Wave phase conjugation is known as an effective tool for compensation of phase distortions arising during wave propagation in inhomogeneous media. Backpropagation of the phase conjugate wave (PCW) towards the emitter restores the phase of the primary wave on its source. This feature of PCW results from reversibility of wave processes in stationary media. The presence of flows or of moving objects in a medium breaks the time-reversal invariance of the wave process and leads, in particular, to the phase shift of the emitted and conjugate waves on the source. In the present paper the theory of PCW propagation in nonlinear and moving inhomogeneous media is developed on the basis of the modified KZK approximation. Applications of the wave phase conjugation for new methods of ultrasonic velocimetry and nonlinear diagnostics of flows are considered.

5pPA9. Ultrasonic scattering in polycrystalline material with elongated grains. Goutam Ghoshal and Joesph A. Turner (Dept. of Eng. Mech., W317.4 Nebraska Hall, Univ. of Nebraska-Lincoln, Lincoln, NE 68588-0526)

The scattering of elastic waves in polycrystalline materials is relevant for ultrasonic materials characterization and nondestructive evaluation (NDE). Ultrasonic attenuation and diffuse backscatter measurements have been especially useful for extracting microstructural information such as grain size and for detecting flaws in materials. Accurate interpretation of experimental data requires robust scattering models. Such models often assume constant density and uniform grain size such that the scattering for grains with mainly spherical geometry is well understood. However, many structural materials are processed in such a way that the microstructure has grain elongation (e.g., from rolling). The scattering of elastic waves in such media must be understood well for design of effective materials inspection methods and for quantitative data analysis. For materials with grain elongation, the appropriate spatial correlation function is no longer

isotropic such that closed-form solutions for attenuation cannot be achieved. Here, attenuation and diffuse backscatter expressions are derived as a function of wave propagation direction with respect to the axis of grain elongation. Numerical results are presented for materials of common interest under assumptions of statistical homogeneity. These results are anticipated to impact ultrasonic nondestructive evaluation of polycrystalline media. [Work supported by U.S. DOE.]

3:45

5pPA10. Accurate time domain computation of linear wave propagation using Chebyshev collocation and matrix diagonalization. Lixi Huang (Dept of Mech. Eng., The Univ. of Hong Kong, Pokfulam Rd., Hong Kong)

Accurate time domain computation of the wave propagation phenomena is of great interest to many fields of studies. The spectral method of Chebyshev collocation is adopted for smooth problems of linear waves. Space differentiation is discretized by Chebyshev derivative matrices and the wave equation is cast as a second-order ordinary differential equation in time domain. The general boundary conditions involving pressure and acoustic particle velocity are discretized and absorbed into the system equations for the unknown vector containing all grid points. The system equations are solved accurately by the method of matrix diagonalization, which involves the finding of the eigenvalues and eigenvectors for the appropriate combinations of the damping and stiffness matrices. For this solution, the time step for stable and accurate computation is only limited by the consideration of round-off errors when the exponential functions are evaluated. Comparison is made with analytical solutions in the one-dimensional case, and good agreement is obtained. The only perceived drawback would be the size of matrices that can be alleviated by the use of domain decomposition approaches. The physical meaning of the eigenvalues derived from the discretization procedure is discussed.

4:00

5pPA11. Ultrasonic characterization of microstructure evolution during processing. Liyong Yang and Joseph A. Turner (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588, jaturner@unl.edu)

Many cold-working processes for polycrystalline metals cause alignment of the grains with a single symmetry axis. This type of microstructure is called fiber texture. The existence of a preferred orientation of the grains has a significant influence on the propagation and scattering of ultrasonic waves, which are often used for material inspection. Knowledge of the wave attenuation of such textured materials is of both theoretical and practical interest to nondestructive testing and materials characterization. In this article, the quantitative relations between fiber texture and wave attenuations of hexagonal crystals are presented. The texture is characterized by a Gaussian distribution function that contains a single parameter that governs the transition of the texture from perfectly aligned crystals to statistically isotropic. Under this assumption, the materials of interest have a varying degree of transverse isotropy representative of processing conditions. Simple expressions for the attenuations of the three modes of waves are given in a concise, generalized representation. Finally, numerical results are presented and discussed in terms of the directional, frequency, and texture dependence. The results presented are expected to improve the understanding of the microstructure evolution during thermomechanical processing. [Work supported by U.S. DOE.]

4:15

5pPA12. Shaped sound focusing using multiple sources: An acoustical ring. Jin Young Park and Yang-Hann Kim (Ctr. for Noise and Vib. Control at KAIST, Japan, jypark1979@kaist.ac.kr)

Shaped sound focusing is defined as the generation of acoustically bright shape in space using multiple sources. The acoustically bright zone is a spatially focused region with relatively high acoustic potential energy level. In view of the energy transfer, acoustical focusing is essential because acoustic energy is very small to use other types of energy. In clinical uses, for example, there are several approaches for using thermal energy

from focused ultrasonic wave energy, so-called high intensity focused ultrasound, concerning point-focusing not regional focusing. If sound shape can be controlled, it offers various kinds of solutions for clinical uses and others because a regional focusing is essentially needed to reduce a treatment time and enhance the performance of transducers. For making the shaped-sound field, control variables, such as a number of sources, excitation frequency, source positioning, etc., should be considered according to geometrical sound shape. To verify these relations between them, a wave-number domain matching method is suggested because the wave-number spectrum can provide the information of the source from the phase and magnitude of the wave-number spectrum. In this presentation, discussions about shaped sound focusing and how to make this sound field are provided for the case of an acoustical ring.

4:30

5pPA13. Noncontact acoustic manipulation using a standing wave field in air. Teruyuki Kozuka, Kyuichi Yasui, Toru Tuziuti, Atsuya Towata, and Yasuo Iida (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), Shimoshidami, Moriyan-ku, Nagoya 463-8560, Japan, kozuka-t@aist.go.jp)

A noncontact manipulation technique is necessary to develop micromachine technology, biotechnology, and new materials processing. The authors have developed an acoustic manipulation technique to transport particles in water [T. Kozuka, Proceedings of WCU2003, Paris, 483–486 (2003)]. Although it is more difficult to generate a strong sound field in air than in a liquid, many researchers are actively studying the trapping of particles, droplets, and aerosols in air using ultrasound. Nevertheless, many of the resultant studies only trap and observe objects; they do not transport them. This paper describes an advanced manipulation technique to transport small objects in air using that scheme. A standing-wave field was generated by two sound beams (40 kHz) using bolted Langevin transducers. The beams' axes cross each other. Expanded polystyrene chips were trapped at nodes of the sound pressure in the crossing sound beams. The trapped position was shifted by changing the phase difference of the two sound beams. In addition, transportation at constant speed of the trapped target was realized by a slight difference in ultrasonic frequency between the two sound beams. This system is also applicable to liquid droplets and aerosols in air.

4:45

5pPA14. Propagation phase and zeros in the complex frequency plane. Yoshinori Takahashi (Kogakuin Univ., Tokyo, Japan, takahashi@acoust.rise.waseda.ac.jp), Mikio Tohyama (Waseda Univ., Tokyo, Japan), and Kazunori Miyoshi (Kogakuin Univ., Tokyo, Japan)

This paper describes the transfer function (TF) including the coherent sound field as the distance between source and observation points increases, from a point of view of the number of zeros on a complex frequency plane. The propagation phase could be estimated according to the linear-regression analysis for narrow frequency-band frequency characteristics of the minimum-phase component [Inst. Electron. Inf. Commun. Eng. Jpn., **J89-A**(4) 291–297 (2006) (in Japanese with English figures)]. This paper analyzes zeros of the TFs for the coherent and reverberant fields based on the modal and random sound field theory. Consequently the distribution of zeros could be estimated in the transition from the coherent to the reverberant conditions.

5:00

5pPA15. Using instabilities in the long wavelength transverse optical mode to generate nuclear fusion. Brian Naranjo and Seth J. Putterman (Phys. Dept., UCLA, 1-129 Knudsen Hall, Los Angeles, CA 90095)

Various crystals such as lithium niobate or lithium tantalate display a phase transition due to an instability in the long wavelength transverse optical mode. These low-temperature phases develop huge electric fields which can accelerate ions to energies where nuclear fusion is observed [Nature **434**, 1115–1117 (2005)]. [Research funded by DARPA, ONR, NSF.]

Session 5pSC

Speech Communication: Production II (Poster Session)

Edward S. Flemming, Cochair

Massachusetts Inst. of Technology, Dept. of Linguistics and Philosophy, 77 Massachusetts Ave., Cambridge, MA 02139-4307

Tokihiko Kaburagi, Cochair

Kyushu Univ., Faculty of Design, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

Contributed Papers

All posters will be on display from 1:00 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 5:00 p.m.

5pSC1. Finite-element method analysis of acoustic characteristics of the vocal tract with the nasal cavity during phonation of Japanese /a/. Hiroki Matsuzaki and Kunitoshi Motoki (Faculty of Eng., Hokkai-Gakuen Univ., S-26 W-11, Chuo-ku, Sapporo 064-0926, Japan)

For this study, the transfer functions and active sound intensities of a vocal tract model with and without a nasal cavity were computed using a three-dimensional finite-element method (FEM). The models were based on vowel data obtained by magnetic resonance imaging (MRI) of the vocal tract with a nasal cavity during phonation of the Japanese /a/. The oral cavity was also coupled with the nasal cavity in a three-dimensional volume of radiation. Effects of wall impedance were also examined. Coupling of the nasal cavity to the oral cavity indicated the following aspects. Additional peaks appeared below 3 kHz for the lossless condition. However, they disappeared in the simulation for the soft wall condition. The sound energy circulation did not occur in the simulation for the soft wall condition. Regarding effects of the wall boundary condition on the spectral envelope, three-dimensional simulations confirmed the upward shift of lower formant frequencies. However, disagreement of the formant frequencies between simulated and real speech should be investigated further by adjusting the wall boundary condition to a more realistic one.

5pSC2. Using tagged cine-MRI and finite-element method to lower bound the number of independently controllable motor units in the tongue. Caroline Essex-Torcaso, William S. Levine (Dept. of Elec. and Comput. Eng., Univ. of Maryland), Maureen Stone, Emi Z. Murano (Univ. of Maryland School of Dentistry, Baltimore, MD 21201), Vijay Parthasarathy, and Jerry L. Prince (Johns Hopkins Univ.)

The biomechanical structure of the tongue is unusual in that there is no rigid structure, such as a bone, for the muscles to act against. In order to understand these biomechanics and their control, a mathematical model has been created. This model has the form of a nonlinear controllable incompressible elastic structure that undergoes large deformations. The model was used to solve a simplified two-dimensional inverse problem. For a simple speech motion (/i-u/), the trajectories of specific points within the tongue were tracked by means of tagged cine MRI. These points correspond to finite element nodes and were used as input to the simplified model. The remaining unknowns in the model are the set of muscle activations that produce the observed motion. Because there is no such set of activations, the set that best approximates the observed motion in a least squares sense was chosen. The results indicate that there must be independently controllable compartments in SL and GG in order to produce this

movement. Three methods were used to improve the representation of the tongue and the resolution of the model: refinement of the number of elements, refinement of the number of muscle activations, and the inclusion of dynamics.

5pSC3. Simulation of phonation control using a tensile vocal-fold model. Chao Tao, Yu Zhang, and Jack J. Jiang (Dept. of Surgery, Div. of Otolaryngol. Head and Neck Surgery, Univ. of Wisconsin Med. School, Madison, WI 53792-7375)

The vocal folds act as an energy transducer that converts aerodynamic power into acoustic power. Therefore, modeling the vocal-fold vibration could provide valuable information for voice synthesis. In previous vocal-fold lumped models, the vocal-fold was simplified as a spring oscillator; the restoring force due to lateral deflection of vocal-fold tissue is represented by equivalent lateral springs. However, the vocal-fold tissue actually contains an intricate extracellular matrix, composed of longitudinal elastin, which lies parallel to the vocal fold. Thyroarytenoid, lateral cricoarytenoid, and cricothyroid muscles control the elongation of these longitudinal elastin and further determine the voice pitch, dysphonia, and rough voice. In this study a tensile vocal-fold model is proposed to simulate vocal-fold vibration. In this model the lateral tissue deflection is represented by lateral spring and the longitudinal elastin elongation is described by nonlinear longitudinal springs. Differential equations are developed to describe the tensile vocal-fold model. The fundamental frequency and phonation threshold pressure are predicted by the proposed vocal-fold model. Numerical simulations prove that the proposed tensile vocal-fold model can accurately describe the muscles control on phonation pitch and so on. [Work supported by NIH.]

5pSC4. A new aerodynamic critical parameter of the speech production: Phonation threshold flow. Jack J. Jiang and Chao Tao (Dept. of Surgery, Div. of Otolaryngol. Head and Neck Surgery, Univ. of Wisconsin Med. School, Madison, WI 53792-7375, jiang@surgery.wisc.edu)

In this paper, the implications of phonation threshold flow (PTF) as an addition to the aerodynamic parameters of speech production system are studied. PTF, the minimum airflow volume velocity able to sustain stable vocal-fold vibration, could have utility in clinical vocal disease assessment just like phonation threshold pressure (PTP) [I. R. Titze, J. Acoust. Soc. Am. **83**, 1536–1552 (1988)]. Furthermore, because glottal airflow can be more easily measured noninvasively than subglottal pressure, PTF could be more suitable for routine clinical assessment than PTP. Theoretical

3371   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3371**

5p SAT. PM

studies indicate that PTF is a sensitive aerodynamic parameter dependent on tissue properties, glottal configuration, and vocal-tract loading. Accordingly, this theory suggests that PTF is reduced by the following: decreasing tissue viscosity, decreasing mucosal wave velocity, increasing vocal-fold thickness, or decreasing prephonatory glottal area. A divergent glottis and low vocal-tract resistance also reduce phonation threshold flow. Finally, the significance of PTF for investigation of the distribution of energy in a vocal-fold vibration system and the potential for application of the PTF parameter are assessed. [Work supported by NIH.]

5pSC5. Lung pressure dependence of glottal sound source.

Hideyuki Nomura and Tetsuo Funada (Div. of Elec. Eng. and Comput. Sci., Kanazawa Univ., Kakuma-machi, Kanazawa-shi 920-1192, Japan, nomu@t.kanazawa-u.ac.jp)

The present study investigates the lung pressure dependence of the vibration of vocal cords using numerical experiments based on our proposed glottal sound source model. The glottal sound source model is described as a coupled problem between unsteady glottal jets and mechanical vocal cords. The vocal cord is assumed to be an elastic cover with effective mass of vocal cord. To simulate the mechanical properties of vocal cords, the elastic cover is supported by distributed small mechanical elements of a spring and damper. The speech production process can be predicted by alternately solving the motion of glottal jets and the vocal cords' vibration. Results of this simulation show that the fundamental frequency of vocal cords' vibration and the propagation velocity of the mucosal wave first increase and then remain constant with lung pressure. The threshold lung pressure of 200–400 Pa and the propagation velocity of mucosal waves is of the order of 1 m/s, which are consistent with measured values. These results suggest that our proposed model is suitable for description of a glottal sound source.

5pSC6. Direct computational method of including piriform fossae and nasal cavity in a time-domain acoustic model of the vocal tract.

Parham Mokhtari (NICT, ATR Cognit. Information Sci. Labs., 2-2 Hikaridai Seikacho, Kyoto 619-0288 Japan, parham@atr.jp)

Frequency-domain simulations of the human vocal tract (VT) have previously shown the importance of including the piriform fossae, which impart a pole and two zeros in the 4–5-kHz frequency range and thereby contribute to speaker individualities. The literature has also shown that time-domain simulation of VT acoustics can result in high-quality synthesis naturally including interactions between the time-varying glottal area and the supraglottal VT. In the present work, the time-domain model of [S.Maeda, Speech Commun. **1**, 199–229 (1982)] was extended to include both left and right piriform fossae as side-branches connected to the main VT, in addition to the nasal tract and sinuses. Departing from Maeda's original implementation owing to the complexity of including more than one side branch, the variables representing acoustic pressure and volume velocity at the piriform fossae and nasal tract junctions were analytically eliminated, and the resulting large system of linear equations were solved simultaneously at each simulation sample. This direct method runs at only a few times real-time on a 1.8-GHz notebook PC, while achieving a more natural sound quality in speech synthesis and control over timbral (or voice quality) features that contribute to each speaker's individuality. [Work supported by NiCT and SCOPE-R of Japan.]

5pSC7. Measures of intraoral pressure pulse shape during stop consonants.

Jorge C. Lucero (Dept. of Mathematics, Univ. Brasilia, Brasilia DF 70910-900, Brazil, lucero@unb.br) and Laura L. Koenig (Haskins Labs, New Haven, CT)

To maintain voicing during a stop consonant, adult speakers typically perform articulatory actions that increase supraglottal volumes and slow the rate at which of intraoral pressure (Pio) rises. In an influential study, Müller and Brown [in *Speech and Language: Advances in Basic Research*

and Practice 4, edited by N. Lass (Academic, New York, 1980), pp. 318–389] proposed a quantitative measure of Pio pulse shape, based on the difference between (a) the slope from pressure at closure to average pressure and (b) from average pressure to pressure at release. We recently applied these techniques to data from women and child speakers of English and observed some shortcomings in Müller and Brown's measurement technique, such that the slope difference angle did not clearly reflect the actual pressure pulse shape. The purpose of this paper is to explore new methods of analyzing intraoral pressure contours, based on fitting a mathematical model to the pressure pulse shapes. It will report our progress and compare our results with the previous technique.

5pSC8. Estimation of area function for three-dimensional vocal tract using finite-element method.

Keita Mochizuki and Takayoshi Nakai (The Grad. School of Electron. Sci. and Technol., Shizuoka Univ., 3-5-1, Johoku, Hamamatsu-shi, Shizuoka, 432-8561 Japan)

Numerous studies have obtained a vocal tract area function using data obtained from magnetic resonance imaging (MRI). However, this approach often results in increased error between formant frequencies derived by recorded voice and estimated frequencies derived by MRI. Two factors engender error: creating the vocal tract model and estimating the area function from the vocal tract model. Generally, the errors from those two factors are not considered independently. This study is intended to establish precise estimation of the area function automatically from a vocal tract model. Pressure contours and sound intensity in the vocal tract are obtained using the finite-element method (FEM); the vocal tract area function is obtained from them. The precision is assessed by comparing the formant frequencies that are derived using the FEM with our method. Results show that the vocal tract area function can be estimated precisely from three-dimensional MRI.

5pSC9. Linear stability analysis of an aeroelastic model of phonation.

Zhaoyan Zhang, Juergen Neubauer, and David Berry (UCLA, 31-24 Rehab Ctr., 1000 Veteran Ave., Los Angeles, CA 90095)

This study presents a linear stability analysis of a two-dimensional aeroelastic model of phonation, which incorporates the interaction between the glottal flow and the vocal folds. The model consists of a wall constriction (shaped by the medial surface of vocal the folds), which is situated in a rigid pipe with an applied potential flow. The vocal-fold constriction is modeled as a plane-strain linear elastic layer. An eigenvalue problem is formulated and the eigenvalues of the coupled system and the eigenshape of the vocal fold surface are calculated. Both linear divergence and flutter instabilities are possible. Phonation onset frequency and pressure are investigated as a function of glottal channel width, vocal fold geometry, and vocal fold material properties. Based on aeroelasticity theory, this model allows the prediction of the effects of changing geometric and viscoelastic properties of the vocal folds on phonation onset.

5pSC10. Comparison of vocal tract shaping patterns derived from articulatory fleshpoint data and MRI-based area functions.

Brad H. Story and Kate Bunton (Dept. of Speech, Lang., and Hearing Sci., Speech Acoust. Lab., Univ. of Arizona, Tucson, AZ 85721)

A recent addition to the University of Wisconsin x-ray microbeam (XRMB) database [Westbury, UW-Madison, 1994] (XRMB) is a speaker from whom MRI-based vocal tract area functions have been previously obtained. The purpose of this study was to compare principal components derived from collections of vocal tract shapes for vowels obtained with these two different techniques. For the XRMB data, midsagittal cross-distance functions representative of each of 11 vowels were first obtained from tongue and jaw pellet positions and the hard palate profile of the speaker. A principal components analysis was then performed on the set of cross-distance functions. Results indicated that the first two orthogonal components (referred to as modes) accounted for more than 90% of the

variance. The shape of these two modes was shown to be similar to those previously determined from the same speakers set of MRI-based area functions. The XRMB-based modes were then used to develop a database-lookup technique where mode coefficients could be determined for time frames throughout a series of vowel sequences. Finally, a procedure is described for using the time-dependent mode coefficients obtained from the XRMB data as input for an area function model of the vocal tract. [Work supported by NIH R01-DC04789.]

5pSC11. Temporal structures of articulatory movements: A contrastive study of stop consonants in Japanese, Korean, and Chinese. Tokihiko Kaburagi and Yoshitaka Nakajima (Faculty of Design, Kyushu Univ., 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540 Japan, kabu@design.kyushu-u.ac.jp)

Stop consonants of Japanese, Korean, and Chinese were studied from an articulatory perspective, particularly in relation to the temporal structure of articulatory movements. In general, acoustical features of stop consonants, such as the amount of aspiration, voice onset time (VOT), and so on, are largely controlled by interarticulatory movements of the tract and glottis. In this study, several speakers were instructed to produce stop consonants of their native tongues. Then, states of the articulators were monitored using an electroglottograph (EGG) and pressure and flow sensors. While the EGG tracked vibratory patterns of the glottis, the pressure and flow sensors were used to measure the air pressure inside the vocal tract and the amount of air flowing out of the mouth. It is noteworthy that these languages have different sets of stops: the voiced and unvoiced stops of Japanese; the aspirated, weak, and strong stops of Korean; and the aspirated and nonaspirated stops of Chinese. Temporal structures of interarticulatory movements were analyzed from the observed data. Their respective acoustical consequences were compared among these languages. [Work supported by JSPS and 21st Century COE, Kyushu University.]

5pSC12. The relation of the temporal variation of F2 to articulator movement. Benjamin V. Tucker (Dept. of Linguist., Univ. of Arizona, P.O. Box 210028, Tucson, AZ 85721-0028, bvt@u.arizona.edu) and Brad H. Story (Univ. of Arizona, Tucson, AZ 85721)

The second formant frequency (F2) has been shown to serve a prominent role in speech perception. The specific relation of the temporal variation of F2 to tongue and lip motion is, however, less well understood. Using the x-ray microbeam database (Westbury *et al.*, 1994), data from six males and six females were extracted for six vowel to vowel (VV) transitions (iu, ia, ua, au, ai, ui). Time-dependent displacements of the lower lip pellet, four tongue pellets, and a jaw pellet were extracted from each of the VV transitions for each speaker. Also extracted for each VV were F1, F2, and F3. Jaw movement was decoupled from the movement of the other fleshpoints following Westbury (1994). Correlations were then calculated between each flesh point and each of the formants on both the x and the y axis of movement. An analysis of the correlation results indicates that F2 was most highly correlated with articulator movement on both the x and y axes.

5pSC13. Processing of Japanese vowel devoicing and the effect of speech rate. Naomi Ogasawara (Dept. of Linguist., Univ. of Arizona, Douglass 200E, P.O. Box 210028, Tucson, AZ 85721-0028, naomi703@email.arizona.edu)

Vowel devoicing (or reduction) in Japanese is an interface between the phonetic and phonological levels (Vance, 1987; Yoshida, 2002). The phonological rule for vowel reduction defines that high vowels (/i, u/) undergo reduction when between voiceless consonants or between a voiceless consonant and a pause. However, there are some cases of vowel reduction that do not fit the rule. A high vowel adjacent to a voiced consonant sometimes becomes reduced, especially in casual speech (Arai, 1999). Even a nonhigh vowel can undergo reduction (Arai, 1999; Vance, 2004). In this study,

the effect of speech rate on the processing of reduced vowels is examined, particularly for vowels that are only likely to be reduced in casual, fast speech. How do listeners recognize vowels that are reduced in fast speech, or the words containing them? A lexical decision experiment showed that listeners found it easier to process fast speech containing reduced vowels than careful speech of isolated words containing them. They also found vowel reduction acceptable for nonhigh vowels in fast speech. This suggests that listeners use knowledge of both phonology and speech rate variability in processing connected speech.

5pSC14. Korean coda cluster simplification: Articulatory study. Minjung Son (Dept. of Linguist., Yale Univ., 370 Temple St., New Haven, CT 06520, and Haskins Labs., 300 George St., #900, New Haven, CT 06511)

Korean exhibits optional coda cluster reduction in lateral-stop (C1C2) sequences. However, this phenomenon has not been tested on the basis of articulatory data. The present study examines articulatory movement using EMMA. For two speakers of Seoul-Korean (e.g., JL and SL), clusters such as /Vlk/ and /Vlp/ followed by /ta/ were elicited at a normal rate along with control tokens (e.g., /Vkta/, /Vpta/, /Vlta/). The stimuli were presented with the relevant word in isolation. Results indicated that neither speaker showed /p/ and /l/ reduction, although there was interspeaker variation in /k/ reduction. While no /k/ reduction occurred in C1C2 for JL, SL demonstrated categorical /k/ reduction in 19% of production, where a longer closure duration of a tongue tip gesture was observed (e.g., possible compensatory lengthening) when it was compared with the control /VltV/; this fails to be addressed in impressionistic transcription and OT analyses [June 11th ICKL, pp. 382–390 (1998); Cho, CLS **35**, 43–57 (1999)]. Lastly, the duration of gestural formation of /k/ in the sequence containing /lk/ was longer than that of /p/ in the sequence containing /lp/, which can be articulatorily supporting evidence to explain more /l/ deletion before /k/ than /p/ in Cho (1999). [Work supported by NIH.]

5pSC15. Intrusive vowels in Cruceno Spanish. Cynthia Kilpatrick, Kathryn McGee (Dept. of Linguist., Univ. of California, San Diego, 9500 Gilman Dr. #0108, La Jolla, CA 92093-0108), and James Kirby (Univ. of Chicago, Chicago, IL 60637)

Intrusive vowels are short vowels appearing within consonant clusters that are not treated phonologically like full vowels. Though reported in many languages, few quantititative studies have examined these vowels. Here, an acoustic study of intrusive vowels in obstruent+tap clusters in the Spanish of Santa Cruz, Bolivia is reported. The results support earlier studies in that the intrusive vowel quality resembles that of the following nucleic vowel, rather than a neutral vowel, and intrusive vowels are significantly longer in clusters with voiced obstruents than in those with voiceless ones. However, results do not fully support all previous phonetic descriptions and related theoretical assumptions. In particular, a significant difference is not found in intrusive vowel length based on variables such as place of articulation of the obstruent, quality of the nucleic vowel, or placement of stress in relation to the cluster. In addition, where previous work finds no significance for position in the word, the present work finds that intrusive vowels are significantly longer in word-medial clusters than in word-initial clusters. The data further suggest that the articulation of obstruent+tap does not require an intrusive vowel, as has previously been claimed, as not all obstruent+tap clusters include an intrusive vowel.

5pSC16. Ultrasound study of velar-vowel coarticulation. Sylvie M. Wodzinski and Stefan A. Frisch (Dept. of CSD, Univ. of South Florida, 4202 E. Fowler Ave., PCD1017, Tampa, FL 33620)

In velar fronting, the closure location for a velar consonant is moved forward along the palate due to vowel context. This study is a replication and extension of a previous study on velar fronting [Wodzinski and Frisch, J. Acoust. Soc. Am. **114**, 2395 (2003)]. In this study, ten participants

produced sentences containing monosyllabic words with /k/ onsets and nine different American English vowels. Ultrasound was used to make measurements of lingual-palatal constriction location at the midpoint of stop closure. Participants were recorded using a head-stabilizing apparatus and an acoustically transparent standoff was used between the ultrasound probe and the jaw. Velar closure location was quantified by the angle of elevation from the horizontal axis of the ultrasound probe to the center of the velar closure. The articulatory frontness of the following vowel was quantified using the frequency of $F2$ at the vowel midpoint. A strong negative correlation between velar closure angle and the following vowel $F2$ was seen for all ten participants. The coarticulatory pattern suggests that each velar-vowel combination results in a distinct closure location for the velar. Implications for models of speech production will be discussed.

5pSC17. Introducing diphthongs to the vowel space. Anna Bogacka (Adam Mickiewicz Univ., Poznan, Poland)

The aim of the study is to provide information on a new vowel space of Polish learners of British English, specifically about acoustic properties of diphthongs. Polish does not have diphthongs, though it has vowel plus glide sequences comparable to English rising diphthongs. Polish has neither vowel duration distinctions nor vowels which are qualitatively identical to the ones appearing in English diphthongs. Studying diphthongs offers the possibility of examining the interplay of substitutions of qualitative and quantitative features. Eight dipthongs were taken into account. The conditioning criteria were quality, duration, occurrence of glottal stops, and degree of nasalization. Twenty subjects were recorded reading 61 sentences with diphthongs embedded in real words. The paper discusses challenges in segmentation. Reported are non-native properties of formant and timing relations and systematic differences in alternations applied to: simple vowels versus diphthongs, rising versus centering diphthongs, and initial versus final phases of diphthongs. Specifically, nasal vocalization was found to depend on the presence or absence of a following fricative, but to be independent of a stressed or unstressed position. The finding is explained by resorting to Polish prosody, which is not stress timed. Suggestions for using the findings in second language instruction will also be offered.

5pSC18. Cues to stop place in stop-liquid clusters. Edward Flemming (Dept. of Linguist. and Philosophy, M.I.T., 77 Massachusetts Ave., 32-D808, Cambridge, MA 02139, flemming@mit.edu)

The acoustic correlates of stop place contrasts in prevocalic position have been studied intensively, but much less is known about stop place contrasts in consonant clusters. An understanding of the nature and distribution of cues to contrasts across a variety of contexts is important to the phonological analysis of the distribution of contrasts. This study extends the characterization of stop place contrasts to stop-liquid clusters /br, dr, gr, bl, gl/, based on acoustic analysis of American English. We examine how well established cues relating to burst duration, burst spectrum, and formant transitions generalize to the stop-liquid context. The relative perceptual weight of burst and formant transitions will be assessed through identification of cross-spliced stimuli.

5pSC19. Acoustic analysis of pressed voice. Carlos Ishi, Hiroshi Ishiguro, and Norihiro Hagita (ATR/IRC Labs., 2-2 Hikaridai, Kyoto 619-0288, Japan)

Pressed voice (''rikimi'' in Japanese) is a voice quality related to the vibratory patterns of the vocal folds. It was recently shown that pressed voice carries important paralinguistic information in Japanese, indicating emotional or attitudinal state of the speaker. In the present work, several acoustic features are investigated, aiming for an appropriate acoustic characterization and an automatic detection of pressed voices. Analysis of pressed voice samples extracted from natural conversational speech firstly shows that irregularity in periodicity (such as in creaky and harsh voices)

is a common but not a strictly determinant feature of pressed voices. Spectral analysis shows that parameters related to spectral slope are effective to identify part of the pressed voice samples, but fail when vowels are nasalized or double-beating occurs in a glottal cycle. Temporal analyses of pressed and nonpressed creaky voices indicate that diplophonia (simultaneous production of two separate tones, when vocal folds oscillate out of phase) frequently occurs in nonpressed creaky voices, while it rarely appears in pressed ones. Further temporal analysis of EGG (electroglottograph) waveforms of acted speech showed the same trends obtained for natural speech, indicating that information about the absence of diplophonia can potentially be used for pressed voice detection.

5pSC20. Determining Bayesian evidence and decay time estimates in acoustically coupled volumes. Tomislav Jasa (Dept. of Elec. Eng., The University of British Columbia, Vancouver, BC, Canada V6T 1Z2) and Ning Xiang (Rensselear Polytechnic Inst., Troy, NY 12180)

Sound energy decay analysis is an important element in understanding acoustical properties of coupled volumes in architectural acoustics. A Bayesian model formulation using Monte Carlo algorithms [Xiang and Goggans, J. Acoust. Soc. Am. **110** 1415–1424 (2001); Jasa and Xiang, *ibid.* **117**, 3707(A) (2005)] has been applied to estimating both the decay times and decay order required in sound energy decay analysis. The need to focus on model selection in a Bayesian model formulation of sound energy decay analysis was presented by Jasa and Xiang [J. Acoust. Soc. Am. **119**, 3343(A) (2006)] along with a discussion of the limitations of the existing Monte Carlo algorithms used for this purpose. This paper will present some recent algorithms that can overcome the limitations of a Monte Carlo approach to model selection and show how these algorithms can be applied to determining both the proper model and decay time estimates for acoustically coupled rooms.

5pSC21. Acoustic analysis and synthesis of laughter. Toshiaki Haga, Masaaki Honda, and Katsuhiko Shirai (Dept. of Sci. and Eng., Univ. of Waseda, 3-4-1 Okubo, Rm. 55S-609, Shinjuku-ku, Tokyo 169-8555, Japan)

Laughter is an important emotional expression in speech. The production is characterized simply by quasiperiodic glottal opening/closing gesture with relatively constant articulatory configuration. The acoustic consequence, however, is complicated and appears in the temporal characteristics of voiced-unvoiced category and pitch, the spectral envelope, and the segmental duration. In this study, acoustic parameters that characterize laughter were investigated from acoustical and perceptual points of view. Laughter /hahaha/ was selected as a typical laughter from conversational speech recording and the acoustic characteristics were compared with those of normally uttered speech /hahaha/. The acoustic analysis was performed in terms of pitch in average and its intrasegmental pattern, the spectral tilt and formant bandwidth, and the segmental duration of the vowel /a/ and the consonant /h/. Then, perceptually significant acoustic features were investigated by using means of replacement of the acoustic parameters of the normally uttered speech by those of the laughter in speech analysis-synthesis system. The results showed that an increased pitch in average and distinct intrasegmental pitch pattern as well as spectral envelope with broad bandwidths are significant features to characterize the laughter. The laughter synthesis by rule was also investigated by manipulating these acoustic parameters.

5pSC22. A study of manual articulatory feature-based transcription of conversational speech. Karen Livescu (Comput. Sci. and Artificial Intelligence Lab., MIT, Cambridge, MA), Xuemin Chi, Lisa Lavoie (MIT, Cambridge, MA), Ari Bezman (Dartmouth College, Hanover, NH), Nash Borges, and Lisa Yung (Johns Hopkins Univ., Baltimore, MD)

This study investigates the manual labeling of speech, and in particular conversational speech, at the articulatory feature level. A detailed transcription, including subtleties such as overlapping or reduced gestures, is useful for studying the great pronunciation variability in conversational

speech. This type of labeling also facilitates the testing of automatic feature classifiers, such as those used in articulatory approaches to automatic speech recognition. For this study, approximately 100 utterances drawn from the SWITCHBOARD database have been transcribed using eight articulatory tiers rather than the traditional single phonetic tier. The tiers include: place and degree for up to two constrictions, nasality, glottal state, lip rounding, and vowel quality. Two transcribers have labeled this set of utterances in a multipass strategy, allowing for correction of errors. Preliminary analysis shows a high degree of intertranscriber agreement. Further analysis of the data is being performed to address a number of questions, such as: How quickly and reliably can this type of transcription be done? What are its advantages and disadvantages relative to purely phone-based transcription? What characteristics of the utterances correspond to high or low transcriber agreement? What can be learned from the data regarding articulatory phenomena such as reduction and asynchrony?

5pSC23. Changes in vocal tract resonance during a pitch cycle. Hironori Takemoto, Tatsuya Kitamura (NICT/ATR Cognit. Information Sci. Labs., 2-2-2 Hikaridai, Seika-cho, Soraku-gun Kyoto, 619-0288, Japan, takemoto@atr.jp), and Seiji Adachi (Fraunhofer Inst. for Bldg. Phys., 70569 Stuttgart, Germany)

Vocal tract (VT) resonances during a pitch cycle were examined using area functions of the five Japanese vowels extracted from MRI data of a male subject. The volume-velocity distribution within the VT at each formant was calculated by a transmission line model, as a function of the glottal area. Our previous study showed that in glottal closure condition the fourth formant was provided by the laryngeal cavity (LC) and the other formants by the VT excluding LC, the VT proper (VTp). During the early stage of glottal opening, the fourth formant disappeared, as it had been provided by the closed-tube resonance of LC. After glottal opening, the single volume-velocity node of the first formant moved to the middle pharynx, shortening the effective VT length and thus increasing its frequency. At the other formants, a node appeared at the junction between LC and VTp, and the VT resonance during the glottal opening condition could therefore be approximated by the VTp. This result indicates that the VTp was responsible for most of the VT resonances throughout a period, while LC provided the fourth formant in the glottal closure period. [Work supported by NICT and MEXT KAKENHI of Japan.]

5pSC24. A study on generation of the voice bar in voiced stops. Jianwu Dang (School of Information Sci., Japan Adv. Inst. of Sci. and Technol., Japan), Takayoshi Nakai (Shizuoka Univ., Japan), Yongnan Piao (Japan Adv. Inst. of Sci. and Technol., Japan), and Shinobu Masaki (ATR, Japan)

The voice bar of the voiced stops is generally considered as a radiation sound from the vibration of the laryngeal wall while the mouth is closed. In this study, we investigated the mechanisms involved in generation of the voice bar of the voiced stops by means of acoustic and mechanical measurements and MRI observations for three Japanese speakers. Radiated sounds from the mouth and the nostrils were recorded separately, while vibration of the laryngeal wall was measured by an accelerometer. The correlation of the nostril radiation to the voice bar is higher than that of the laryngeal wall vibration to the voice bar. It implies that the major component of the voice bar is likely radiated from the nostrils but not the laryngeal wall, where the nostrils' radiation is possibly caused by a trans-velar coupling of the yielding velum. The authors have proposed a two-layer board model to imitate the transvelar function. To obtain physiological evidences, articulation of stop-vowel sequences was observed by MRI movies. It found that the thickness of the velum varies with the adjacent vowels. Applying this measurement to the two-layer board model, the generation of the voice bar was evaluated by comparing the simulations and observations.

5pSC25. Sound radiations from nostrils for voiced speech and their individuality. Takayoshi Nakai and Masashi Yamamoto (Faculty of Eng., Shizuoka Univ., 3-5-1 Johoku, Hamamatsu, Japan, 432-8561tdtnaka@ipc.shizuoka.ac.jp)

This study is intended research to analyze sound radiations from the nostrils for voiced speech and explore their characteristics. Sound pressure near the nostril, vibration acceleration of the nose, and sound pressure in front of the mouth were observed using a three-channel A-D converter. The ratio of radiations from the nostril and the mouth, RRNM, was observed. Speech materials included nasal sounds /ma/ and /na/; five vowels; and /cvcv/(c=b, d, g, v=a, i, u, e, and o). There are ten male subjects. For initial buzzes, RRNM showed individuality: it is from 0 to −15 dB lower than for /m/. For medial buzzes, RRNM is mostly lower than for initial buzzes. Transitions from buzz to vowel and from vowel to buzz were mostly fast, but their transitions are slow and smooth in some cases. Frequency analyses revealed that radiation from the nostrils consists mainly of fundamental frequency, and the second and third harmonics.

5pSC26. Spectral characteristics of period doubled phonation. Jody Kreiman and Bruce R. Gerratt (Div. of Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, jkreiman@ucla.edu)

Period doubling (also called subharmonic phonation, bifurcated phonation, F0 series) occurs frequently in both pathological and normal phonation. In this kind of phonation, the most obvious acoustic characteristic is that vocal pulses appear in pairs, with alternate pulses differing in period and often in amplitude. Attempts at modeling period doubling by varying period and amplitude did not capture its characteristic quality, leaving the question of which acoustic variables underlie our perception of period doubling. Acoustic analyses of 50 period doubled voices were undertaken to determine what acoustic features might underlie the constant percept associated with period doubling. Every voice examined did show alternation between short/long and large/small glottal pulses. However, alternations in amplitude were not consistently linked to alternations in period. Voices additionally showed a consistent pattern of alterations in pulse shape, so that the A and B pulses consistently differed in H1-H2. Results suggest that the consistent vocal quality associated with period doubling is due to combined alternations in period and spectral tilt. Amplitude alternations appear to be acoustic artifacts of the period and spectral alternations.

5pSC27. On the vocalization of nonspeech sounds: Implicit mechanisms. Rodrigo Segnini (Shimojo Implicit Brain Function Project, JST-ERATO, 3-1 Morinosatowakamiya, Atsugi, Kanagawa, Japan 243-0198, rsegnini@shimojo.jst.go.jp) and Takemi Mochida (NTT Sensory and Motor Res. Group, Atsugi, Kanagawa, Japan 243-0198)

When humans vocalize nonspeech sounds, as when imitating other sound-producing mechanisms, how effective is this vocalization? The human vocal tract is adept at producing speech sounds, allowing verbal communication despite variations in these sounds due to differences in size, shape, and dynamics of this mechanism among individuals. In contrast, nonspeech sounds that were not originally produced by the human vocal tract derive from even more structurally different mechanisms—those of animal calls or machinery sounds, for example. Through experiment, we determined that in imitative tasks of such nonspeech sounds, subjects were intuitively able to produce distinguishing articulatory movements for each sound under different conditions, even when lacking auditory feedback. These results reveal a nontrivial coherence between articulatory events and the intended sound despite the noted differences with the original production mechanism and the resulting acoustic mismatch. The authors would like to express sincere appreciation to Makio Kashino for enriching discussion and support in the realization of this project.

5p SAT. PM

5pSC28. Basic processes underlying speaking and singing: Preliminary acoustic and aerodynamic comparisons. Amee P. Shah (Dept. of Health Sci., Cleveland State Univ., 2121 Euclid Ave., MC 431-B, Cleveland, OH 44115)

Intuitively, we think speech and singing are different. Perceptually, we hear the difference. Aesthetically, we appreciate the difference. Psychologically, we benefit from the positive effect of singing. The important question is whether this perceived difference is purely perceptual or is it grounded in physiological reality? It is important to identify the difference between speaking and singing if indeed there is one as it can shed light on our theoretical understanding of normal auditory perception and vocal production processes. Addressing this question has practical applications as well, as the findings could yield valuable information for working with clients in the fields of speech pathology and vocal music training. In addressing this question, the present investigation targeted broad areas to identify the differences between speech and singing. Specifically, respiration (the mechanics of breathing), articulation (the movements of the oral musculature), phonation (vocal-fold movement and air pressures), and acoustic attributes (temporal and spectral measures) were compared across speaking and singing. Surprisingly, and somewhat counterintuitively, results showed that speaking and singing are not as different as they are perceived to be. Results of preliminary acoustic and aerodynamic comparisons will be discussed as they help better understand the processes of singing and speaking.

5pSC29. The articulation of consonants in Kinyarwanda's sibilant harmony. Rachel Walker, Dani Byrd, Sungbok Lee, Celeste DeFreitas (Dept. of Linguist., USC, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, rwalker@usc.edu), and Fidele Mpiranya (Univ. of Chicago, Chicago, IL 60637)

Kinyarwanda's sibilant harmony causes alveolar /s, z/ to become retroflex when preceding a retroflex fricative within a stem. Intervening coronal stops block sibilant harmony, but bilabial and velar consonants are transparent. This study investigates the production of Kinyarwanda's sibilant fricatives and also examines transparent and opaque consonants in the system. Articulatory kinematic data were collected for a native speaker of Kinyarwanda using electromagnetic articulography. This allowed calculation of the mean angle for receivers affixed to the tongue tip and blade over the target consonant intervals. Average mean angle reliably and robustly distinguished alveolar and retroflex fricatives, with alveolars showing a lower tip relative to blade. No significant difference in mean angle was found for /t/ in contexts where it blocked retroflex sibilant harmony versus ones where it preceded an alveolar fricative, confirming that /t/ does not participate in the harmony. However, in contexts where /m/ and /k/ are perceived as transparent to sibilant harmony, their mean tip-blade angle was significantly different from contexts where harmony did not occur. Furthermore, mean angle during transparent /m/ showed no significant difference from mean angle during retroflex fricatives, suggesting that tongue tip-blade angle conducts strongly and systematically through transparent consonants but without perceptible effect. [Work supported by NIH, USC URP.]

5pSC30. Validity of the Nasometer measuring the temporal characteristics of nasalization. Youkyung Bae, David P. Kuehn, and Seunghee Ha (Dept. of Speech and Hearing Sci., Univ. of Illinois at UC, 901 S. Sixth St., Champaign, IL 61820, yousong@uiuc.edu)

Patients with oral-nasal resonance imbalance may present different characteristics not only in the amplitude but also in the temporal aspects of nasalization. Although nasalance has been the most commonly derived measure from the Nasometer (KayPENTAX), the Nasometer also might be useful in providing another important measure, time. Such measurements, however, should be validated in relation to a separate external criterion procedure. This study examined the validity of the Nasometer measuring the temporal characteristics of nasalization. Fourteen adult American English speakers produced three rate-controlled nonsense speech samples:

/izinizi/, /azanaza/, and /uzunuzu/. Acoustic data recorded through the Nasometer were compared to those recorded through a standard audio recording setup which served as an external criterion procedure. Four timing variables pertinent to nasalization were measured from the digitized acoustic signals: nasal onset, nasal consonant, nasal offset, and total nasalization. No significant differences were found between the two instrumental conditions in any of four timing variables measured, which suggests the Nasometer can be used as a valid tool to measure the temporal features of nasalization. This study also provided a valid way of segmenting speech tasks using the Nasometer and confirmed the significant effect of different vowel contexts on the temporal characteristics of nasalization.

5pSC31. Pitch determination and sinusoidal modeling for time-varying voiced speech. Masashi Ito and Masafumi Yano (Res. Inst. of Elec. Commun., Tohoku Univ., 2-1-1 Katahira. Aoba-ku, Sendai 980-8577, Japan, itojin@riec.tohoku.ac.jp)

A voiced speech signal consists of sinusoidal components of which amplitude and frequency are time varying. Usually the signals are analyzed by assuming that they are stationary within a local analysis window, so they include inevitable errors. To solve this problem, we have proposed a method termed local vector transform (LVT), in which amplitude and phase of the sinusoidal components were approximated by quadratic functions uniquely determined from input spectrum. Two types of experiments were carried out to evaluate the validity of this method. First, time-varying pitch frequencies (F0s) of natural utterances were investigated. F0 determined by LVT greatly improved the accuracy compared to the conventional pitch determination algorithms: cepstrum, autocorrelation, and instantaneous frequency estimation. Second, LVT is applied to the sinusoidal modeling and the amplitude and phase were estimated for every component. The results apparently showed that the signal obtained by LVT was very close to the input. This was quantified by a signal to residual power ratio (S/R). For both of synthesized and naturally uttered speech signals, LVT showed the higher S/R compared to the conventional algorithm. These results indicate that the proposed algorithm is highly effective for pitch determination and sinusoidal modeling for time-varying speech signals.

5pSC32. The effect of stress on consonant vowel (CV) coarticulation: Decoupling of the CV bond. Augustine Agwuele, Harvey Sussman, and Bjorn Lindblom (Dept. of Linguist., Univ. of Texas, Austin, TX 78712)

Two common sources of variation in the speech signal arise from (1) naturally imposed speaking conditions (e.g., stress, tempo), overlaid onto (2) inherent contextual influences. Coarticulatory analyses are affected by both sources of variation, but rarely are they dissociated. The focus of this study was to derive quantitative methods to assess the effect of emphatic vowel stress on preceding stop consonant onsets, apart from, and independent of, anticipatory effects of changing vowel contexts. Three speakers produced V1CV2 tokens, with emphatic stress either on V1 or V2; C =/bdg/. There were six V1 and 10 V2 vowel contexts. Novel methods were applied to both standard locus equation scatterplots and multiple regression analyses to isolate F2 midvowel (Hz) and F2 onset (Hz) adjustments as a function of emphatic stress. The analyses show how stress differentially affects the V-midpoints and C-onsets for labial /b/ relative to lingual stops /dg/.

5pSC33. Intrinsic vowel F0 and speech rate. Ian Maddieson (Univ. of New Mexico/Univ. of California, Berkeley Dept. of Linguist., Univ. of New Mexico, MSC 03 2130, Albuquerque, NM 87131)

A relation between vowel height and F0 such that, other things being equal, F0 is a little higher in a high vowel than a low vowel appears to be a phonetic universal. However, the cause of this relationship remains unclear. An insight might be gained from ways in which the relationship changes under transformations of speech rate. Ten speakers of American

English were recorded saying sentences containing words such as "bead" and "bad" at fast and slow rates. For the majority of the subjects the intrinsic F0 difference was greater in fast speech, contrary to expectation. The difference cannot be explained by changes in vowel quality between the conditions.

5pSC34. Tongue dorsum location and tongue root retraction in alveolar and palatal clicks in the endangered language N/uu. Amanda L. Miller, Johanna Brugman, Jonathan Howell (Dept. of Linguist., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853), and Bonny Sands (Northern Arizona Univ., AZ)

Traill (1985) describes Xoo clicks as having velar posterior constriction locations (PCL). Miller *et al.* (to appear) show that the PCLs of Khoekhoe alveolar and palatal clicks are uvular and uvulo-pharyngeal. Ladefoged and Maddieson (1996) discuss clicks involving a uvular posterior release location (PRL) (and an implied velar PCL). An ultrasound investigation of velar and uvular pulmonic stops, alveolar and palatal velar clicks, and palatal uvular clicks, in the endangered language N/uu is presented. Results are for 15 repetitions of each consonant in the u context by four of the remaining speakers. The alveolar click has the tongue dorsum (TD) between the velar and uvular stops, and tongue root retraction (TRR) like the uvulars. The TD in the palatal click is also similar to the velar and uvular stops, but there is no TRR. The TD in the uvular palatal click is at the uvulo-phraryngeal location, and there is TRR similar to the uvular stop, in both the closure and the release. TRR in the alveolar-uvular and palatal-uvulo-pharyngeal clicks explains their phonological patterning with uvulars. The lack of TR retraction in the palatal click explains its patterning with velars. [Work supported by NSF, BCS #0236735.]

5pSC35. Acoustic evidence for the lenis/fortis contrast in California Northern Paiute. Molly Babel (UC Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650, mbabel@berkeley.edu)

California Northern Paiute (CNP) is spoken in the communities on the eastern slopes of the Sierra Nevada Mountains in northern California. CNP is the southernmost dialect of Northern Paiute, a language belonging to the Western Numic branch of Uto-Aztecan. CNP is a severely endangered language spoken by fewer than ten fluent speakers. CNP has a phonemic contrast that is distinct from that of any other Numic language. CNP has three series of stops: lenis, fortis, and voiced fortis. Previous descriptions of CNP have used only impressionistic judgments to categorize the phoneme series. This paper presents the first acoustic evidence for the lenis/fortis contrast in CNP. Field recordings from five native speakers of CNP were analyzed in this investigation. The lenis/fortis contrast was examined by measuring the closure duration using cues from the waveform and spectrogram in a digital signal processing program. Results of duration measurements support the impressionistic evidence and suggest a voiced and voiceless fortis series in addition to a single lenis series in the phonemic inventory of CNP. This research has implications for the acquisition of categories by second language learners who are first language speakers of a language that does not have such a contrast.

5pSC36. Articulation without acoustics: A case study of Oneida utterance-final forms. Bryan Gick (Dept. of Linguist., University of British Columbia, E270-1866 Main Mall, Vancouver, BC V6T 1Z1, Canada), Karin Michelson (Univ. at Buffalo, Buffalo, NY 14260), and Bosko Radanov (University of British Columbia, Vancouver, BC V6T 1Z3, Canada)

Whether the targets of speech production are primarily articulatory or acoustic has been controversial [Guenther *et al.*, J. Acoust. Soc. Am. **105**(5), 2854–2865 (1999)]. While it has been observed that tongue movements occur even in contexts where they may be acoustically obscured [Browman and Goldstein, LabPhonI: 341–376 (1990); Tiede *et al.*, J. Acoust. Soc. Am. **110**(5), 2, 2657 (2001)], the question of whether natural languages can systematically encode articulations in the absence of acoustic consequences has remained open. A study was conducted to investigate a purported phonological process in the endangered Oneida (Iroquois) language whereby utterance-final forms exhibit fixed sequences having this property. Ultrasound, video, and acoustic data were collected from two native speakers of Oneida in a field setting. Preliminary results indicate that speakers' tongue positions are significantly different during different utterance-final vowel articulations despite these vowels being completely inaudible. Results confirm claims that the Oneida phenomenon is in fact a case of articulation without acoustics, and indicate that it is indeed possible for articulations to be retained even in the systematic absence of auditory goals. [Work supported by NSERC.]

5pSC37. Automatic detection of vowel nasalization using knowledge-based acoustic parameters. Tarun Pruthi and Carol Y. Espy-Wilson (Dept. of Elec. Eng. and Inst. of Systems Res., Univ. of Maryland, College Park, MD 20742, tpruthi@umd.edu)

Researchers in the past have suggested several acoustic correlates of nasalization including extra pole-zero pairs near the first formant ($F1$), a reduction in $F1$ amplitude, and an increase in F1 bandwidth. Even though these correlates have been known for a long time, considerable work is still needed to automate the extraction of acoustic parameters (APs) for nasalization. This work looked at 37 different APs which were pared down to 8 APs based on F statistic obtained by ANOVA. In preliminary experiments, an accuracy of 69.79% has been obtained for the task of discriminating between oral and nasalized vowels on the TIMIT database using a support vector Machine (SVM)-based classifier. The classification was done on a frame basis, and a segment was declared nasalized if more than 30% of the frames were found to be nasalized. Note that all vowels adjacent to nasal consonants were assumed to be nasalized. Thus, the accuracy may actually be higher since some vowels before nasal consonants may not be nasalized. Further, these results were obtained by using a linear kernel in SVMs. We hope the results would improve when a radial basis function kernel is used. [Work supported by Honda and NSF Grant BCS0236707.]

5pSC38. A novel method for the measurement of vowel formant frequencies and its application to the analysis of Japanese vowels over a wide range of ages. Hideki Kasuya (Dept. of Speech, Lang. and Hearing Sci., Intl. Univ. of Health and Welfare, Otawara, Japan 324-8501, kasuyah@snow.ucatv.ne.jp), Kanae Ohta, Hiroki Mori (Utsu-nomiya Univ., Utsunomiya, Japan 321-8585), Toshisada Deguchi, Ghen Ohyama (Tokyo Gakugei Univ., Tokyo, Japan 184-8501), and Yukimasa Muraishi (Univ. of Tokyo, Tokyo, Japan 164-8654)

The first three formant frequencies of the Japanese vowels, measured over a wide range of ages of talkers by one of the authors in 1968, have widely been cited in many speech textbooks and research papers. We decided to update those old data, measured by sound spectrograms, by using a more advanced method. We developed an iterative method for measuring vowel formant frequencies based on an autoregressive with exogenous input (ARX) speech production model that incorporates both voice source and vocal-tract models [J. Acoust. Soc. Jpn. **58**, 386–397(2002)]. In using this method to measure vowel formant frequencies of high-pitched voice, the procedure for determining the initial value of the open quotient (OQ) of the source model was carefully investigated, since it affects significantly the estimation of the lower first formant frequency. Estimated OQs of the vowel /a/ of a talker were found to be stable against various initial values and so were used for analyzing the other vowels of the same talker. Perceptual confirmation of phonetic quality of resynthesized vowels with the fundamental frequency perturbed was also of great importance for accurate measurement. This procedure was successfully applied to 320 male and female talkers, aged 6–24 years.

5pSC39. Spectral properties of Japanese whispered vowels referred to pitch. Hideaki Konno, Hideo Kanemitsu (Hokkaido Univ. of Education, Hakodate 040-8567, Japan, konno@cc.hokkyodai.ac.jp), Jun Toyama, and Masaru Shimbo (Hokkaido Info. Univ., Ebetsu 069-8585, Japan)

Whispered speech can communicate the same linguistic information as ordinary speech in spite of the great difference of their respective acoustic characteristics. In this respect, whispering is an interesting object for studying speech perception and recognition. In this study, we investigate the spectral properties of five whispered Japanese vowels uttered in isolation and having different pitch. Pitch of ordinary whispered vowels was measured in terms of the manner in which the talkers were listening to pure tones while uttering and adjusting its frequency so that the pitch matched the utterance. For other samples, talkers changed the pitch of utterances to match the given pure tones. Acoustic analyses were carried out on formant frequencies, a spectral tilt, and a peak frequency of wide-ranging spectral shape using second-order LPC method called a global peak. Preliminary results show a tendency of upward shift of $F1$, $F2$, and global peaks, and flattened spectral tilts on overall vowels with increasing pitch. The intended pitch by a talker is not correspondent to a specific formant frequency. Results of pitch comparison tests among samples and pitch matching test with pure tones will also be discussed. [Work supported by MEXT Japan.]

5pSC40. Detecting nonlinearity in voice production systems. Yu Zhang, Chao Tao, and Jack J. Jiang (Dept. of Surgery, Div. of Otolaryngol. Head and Neck Surgery, Univ. of Wisconsin Med. School, Madison, WI 53792-7375)

Many biomedical systems exhibit some form of nonlinear behavior, such as saturation in response to certain inputs. Linear description may not be sufficient to describe the complex nonlinear phenomena, including bifurcation and chaos. The determination of nonlinearity from time series has become an important topic for the analysis and modeling of biomedical systems. In this study, we will apply nonlinear dynamic modeling based on Volterra kernel function to describe the nonlinearity of voice signals. Three nonlinear parameters including nonlinearity-to-linearity amplitude ratio (NLAR), nonlinearity-to-linearity parameter ratio (NLPR), and nonlinear order are employed. Sufficiently lower NLAR and NLPR, and a nonlinear order approaching 1 can be found in periodic voice signals, demonstrating the linear mechanism of periodic voice production. However, aperiodic voices show sufficiently high NLAR and NLAR, and a nonlinear order above 2, indicating the dominant role of nonlinearity in disordered voice production. Furthermore, the effects of the signal length and noises on these three parameters are investigated. These three nonlinear parameters show the robustness of short signal length and noise perturbations, demonstrating their potential applications in measuring the nonlinearity of disordered voice production systems. [Work supported by NIH.]

5pSC41. Concurrent speech disturbs word generation: Semantic, associative, and grammatical processes in picture naming: A pictureword interference study. Masumi Watanabe (Dept. of SLP, Tama Rehabilitation Acad., 1-642-1 Nekabu, Ome, Tokyo, Japan 198-0004), Kazuhiko Kakehi (Cyukyo Univ., Kaizu-cho, Toyota, Japan, 470-0393), Joanne Arciuli, David Vinson, Gabriella Vigliocco (Univ. College London, London WC1H 0AP, UK), and Noriko Iwasaki (UC Davis, Davis, CA 95616)

In pictureword interference (PWI), a picture of an object to be named is presented with a distracter word. Most PWI studies showed semantic interference between a target picture and distracter word (noun). Recently, Vigliocco et al. (2005) studied grammatical effect in naming a picture of action. They found that generation of an inflected form of an Italian verb was disturbed by a distracter in the same grammatical class. In experiment 1 of the present study, naming an object with an auditorily presented noun or verb distracter in Japanese was investigated, and grammatical class but not semantic effect was found, which is incongruent with previous find-

ings. Analyzing the distracters revealed that familiarity of the verbs was lower than that of nouns, and half of nouns classified as semantically close were associative (e.g., garage versus car). In experiment 2, familiarity was matched across distracters, and association effect was examined separately. To see time course of interference, stimulus onset asynchrony (SOA) was changed from -300 to $+200$ ms. Again, grammatical effect was observed, and associative as well as semantic effect was found for negative SOAs. Possible factors producing grammatical and associative effects in PWI will be discussed cross-linguistically.

5pSC42. Quantifying the Lombard effect in different background noises. Christian Giguère, Chantal Laroche, Émilie Brault, Julie-Catherine Ste-Marie, Marianne Brosseau-Villeneuve, Bertrand Philippon, and Véronique Vaillancourt (Univ. of Ottawa, 451 Smyth Rd., Ottawa, ON, Canada K1H 8M5)

The Lombard effect of increasing one's vocal effort in the presence of background noise has been quantified by Pearsons et al. (1977) as a 0.6-dB increase in speech levels for each dB increase in the background noise up to a ceiling level. Lombard speech has also been investigated in other studies with variable results. This study reports data on the effect of different noises on (1) the slope of the function relating speech levels and noise levels and (2) the spectral structure of speech. Twenty normal-hearing adults were asked to read aloud ten sentences from the hearing in noise test (HINT) to an experimenter seated 1 m away, in quiet and in various noises (white, speech spectrum, babble, and restaurant) presented in the sound field at 60 and 75 dBA. Preliminary findings show that increases in speech levels in natural environmental noises (restaurant and babble) most closely follow Pearsons' data, with a slope of 0.6 dB. In contrast, artificial noises (speech spectrum and white) were associated with lower slopes (0.2 and 0.4 dB, respectively). Findings of this study could be useful in a wider context of modeling the complete speech communication process from talker to listener.

5pSC43. Frequency modulated feedback of ones own voice can cause dysfluence of speech. Xu Yizhong (Res. Inst. for Electron. Sci., Hokkaido Univ., N12-W6 Kita-ku, Sapporo 060-0812, Japan), Akira Toyomura, Sachiko Koyama, and Shinya Kuriki (Hokkaido Univ., Sapporo 060-0812, Japan)

Delayed auditory feedback (DAF) can cause speech dysfluency (e.g., stuttering) in healthy normal subjects. In previous studies, feedback frequency was changed linearly or to a constant value in FAF. In this study, we introduce frequency-modulated feedback of pitch in a sine-wave manner. The modulation depth of pitch ($F0$) of auditory feedback voice was six semitones, and the modulation frequencies of sinewaves were set at 0.05, 0.1, 0.5, 0.9, 2, 4, 6, 8, 10, 12, 14, and 16 Hz. In addition to the FAF experiment, we conducted DAF experiments with delay times of 50, 200, and 400 ms. Participants (18 Chinese and 18 Japanese) were instructed to read sentences as they listened to the altered feedback voice through a headphone under each condition. Speech dysfluency was apparent under all FAF and DAF conditions for both groups. Only when the modulation frequency was 14 Hz did Chinese subjects show significantly larger disturbance than Japanese in FAF. A significant greater disturbance of Chinese in DAF was observed when the delay time was 50 ms. These results indicate that frequency modulation can cause speech dysfluency, as does DAF. [Work supported by the 21st century COE (Center of Excellence) Program.]

5pSC44. An online customizable music retrieval system with a spoken dialogue interface. Sunao Hara, Chiyomi Miyajima, Katsunobu Itou,[a] and Kazuya Takeda (Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan, hara@sp.m.is.nagoya-u.ac.jp)

In this paper, we introduce a spoken language interface for music information retrieval. In response to voice commands, the system searches for a song through an internet music shop or a "playlist" stored in the local PC; the system then plays it. To cope with the almost unlimited size

of the vocabulary, a remote server program with which a user can customize their recognition grammar and dictionary is implemented. When a user selects favorite artists, the server program automatically generates a minimal set of recognition grammars and a dictionary. The system then sends them to the interface program. Therefore, on average, the vocabulary is less than 1000 words for each user. To perform a field test of the system, we implemented a speech collection capability, whereby speech utterances are compressed in free lossless audio codec (FLAC) format and are sent back to the server program with dialogue logs. Currently, the system is available to the public for experimental use. More than 100 users are involved in field testing. In our presentation, we will report details of the system and the results of field tests, which include motorcycle environments. [a]Currently at Graduate School of Computer and Information Sciences, Hosei University, Tokyo 184-8584, Japan.

5pSC45. Database construction and analysis of user speech with real environment spoken guidance systems. Hiromichi Kawanami, Manabu Kida, Tobias Cincarek, and Kiyohiro Shikano (Grad. School of Information Sci., Nara Inst. of Sci. and Technol., 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101, Japan, kawanami@is.naist.jp)

A spoken guidance information system "Takemaru-kun" has developed and been operated in a public city center in Ikoma city, Nara, Japan since November 2002. The system employs the example-based response selection strategy and answers to users' questions about the hall facility, the city, sightseeing, etc. with CG agent animation. Following the success of the system, two new spoken systems "Kita-chan" and "Kita-chan robot" were developed and were set on a railway station since March 2006. "Kita-chan robot" is being operated with a robotlike agent. All user speech of the systems is recorded and two year's data of "Takemaru-kun" and one month's data of "Kita-chan" systems are manually transcribed and labeled by age and gender. Proper answers are also given to each utterance. In this paper, details of the three user speech databases are described and analytical results concerning numbers of inputs and response accuracy and age group and gender are reported.

5pSC46. A speech communication environment using open source software library for active sound image control. Yuichiro Kitashima, Kazuhiro Kondo, and Kiyoshi Nakagawa (Yamagata Univ., Jounan 4-3-16, Yonezawa, 992-8510, Yamagata, Japan)

This paper describes a three-dimensional (3-D) conference system using an open source software library on conventional PCs. We will attempt to use both 3-D graphics and audio to construct a virtual conference environment for effective communication between remote parties. A rough prototype system was developed using OpenGL and OpenAL. The system uses local files for voice output, whose image location is rendered according to the user input. We initially evaluated the perceived sound image location accuracy of the rendered sound image using the prototype. Users were asked to identify the location of the rendered sound image from among four choices: front, back, left, and right. The users were able to identify the left and right images correctly at virtually 100%, but the front and back identification were lower than 10% for some sounds, particularly male speech. We plan to implement audio-streaming functions to achieve real-time audio conferencing and evaluate the benefits of 3-D audio for conferences. We would also like to implement HRTF (head-related transfer function) and RTF (room transfer function) for improved audio image localization, especially to achieve accurate elevation perception.

5pSC47. Khmu voice register: Acoustic analysis and perceptual experiments. Arthur S. Abramson, Patrick W. Nye (Haskins Labs., 300 George St., New Haven, CT 06511, arthur.abramson@uconn.edu), and Theraphan L-Thongkum (Chulalongkorn Univ., Bangkok 10330, Thailand)

Khmu has phonological voice registers, i.e., bundles of phonetic properties pertinent to the syllable. Auditory observations claim clear voice and high pitch for register 1 and breathy voice and low pitch for register 2. Although Suai registers were in flux [Abramson *et al.*, Phonetica **61**, 147–171 (2004)], it was understood that Khmu Rawk had a stable distinction. Words were recorded by 25 native speakers. Acoustic analysis yielded F0 and overall amplitude contours, frequencies of F1 and F2 in quasi-steady states of the vowels, relative intensities of higher harmonics to that of the first harmonic, and vowel durations. When circumstances caused early attention to perception testing, the words of only eight speakers had been analyzed. Since the only significant factor that had emerged by then was F0 contour, the synthetic stimuli were made with a series of seven contours. The labeling by 32 native speakers yielded two categories, demonstrating the sufficiency of F0 as an acoustic cue. The completed acoustic analysis showed a significant effect of the harmonic ratios for the women, suggesting a conservative bias. The language is shifting toward tonality. Further perception testing must be done for phonation-type effects. [Work supported by NIH and the Thailand Research Fund.]

Session 5pUW

Underwater Acoustics and Acoustical Oceanography: Scattering, Reverberation, and Noise

Hiroyuki Hachiya, Cochair

Chiba Univ., Research Ctr. for Frontier Medical Engineering, 1-33 Yayoi-cho, Inake-ku, Chiba 263-8522, Japan

Dezhang Chu, Cochair

Woods Hole Oceanographic Inst., Dept. of Applied Ocean Physics and Engineering, Woods Hole, MA 02543

Chair's Introduction—1:00

Contributed Papers

1:05

5pUW1. Modeling scattering from simple three-dimensional bathymetric features using wave-field superposition. John A. Fawcett (DRDC Atlantic, P.O. Box 1012, Dartmouth, N.S., Canada, B2Y 3Z7, john.fawcett@drdc-rddc.gc.ca)

In the method of wave-field superposition a set of point sources inside or outside of a scattering region can be used to represent the exterior or interior acoustic fields. A method is presented that represents a simple bathymetric feature as a closed scattering region in an otherwise flat water-seabed environment. In particular, the feature is taken to be azimuthally symmetric about the vertical axis, thus allowing the three-dimensional scattering problem to be solved in terms of a set of two-dimensional problems. These problems are solved using the method of wave-field superposition. Numerical examples are presented. As well, a benchmark solution in the case that the surrounding medium has only a density jump at the seabed is derived and compared with the more general numerical approach.

1:20

5pUW2. Finite-difference modeling of scattering from shallow buried objects in sea floor. Inger Helene Moen and Hefeng Dong (Acoust. Group, Dept. of Electron. and Telecommunications, NTNU, O.S. Bragstads plass 2B, 7091 Trondheim, Norway)

An object buried in the sea floor positioned close to the water/sediment surface can, when exposed to an incident wave, be a source of interface wave generation. The amplitude of a surface wave tends to dominate at greater distances, and surface waves can thus be a viable tool for detecting shallow buried objects. In this paper, the 2-D velocity-stress finite-difference formulation of Virieux [Geophysics **51**, 889–901 (1986)] is adopted to model the wave propagation from a point source located above the sea floor in an open ocean environment. The wave responses from an object situated below the surface are computed and analyzed, with respect to various object depths, properties, and sizes. A second consideration of the paper is the analysis of the artificial reflections from the model boundaries. The perfectly matched layer absorbing boundaries is a widely used technique that has been shown to be very efficient and has therefore been implemented in order to reduce the grid reflections. The numerical results show that the amplitude of the surface wave decreases with the depth of the object, and that interface waves can be of great importance for detection and identification of shallow buried objects in the sea floor.

1:35

5pUW3. Time-frequency analysis of backscattered echoes from absorbing spherical shells. Hui Ou, Xudong Wang, Vassilis L. Syrmos (Dept. of Elec. Eng., Univ. of Hawaii at Manoa, Honolulu, HI 96822), and John S. AllenIII (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

The backscattering of tone bursts by spherical shells immersed in water has been investigated by several researchers. For a very thin shell, the backscattering specular echo is followed by enhanced echoes. The mechanism of these enhanced echoes has been attributed to a subsonic surface guided wave from recent ray synthesis analysis [P. L. Marston and N. H. Sun, J. Acoust. Soc. Am. **96**, 1862–1874 (1992)]. In this study, time-frequency analysis is performed on the scattering signal. A discrete Wigner distribution function is employed to express the information content in the scattered sound in relation to the size and material of the shell [G. C. Gaunaurd and H. C. Strifors, IEEE **84**, 1231–1248 (1996)]. The simulation results support that the time delay between the echoes is determined by the outer radius of the shell as well as the shell material (with or without absorption), or the Rayleigh velocity of the shell material. Furthermore, the backscattered enhancement occurs at different ranges in frequency domain due to various shell materials, which provide a potential method to determine the shell material properties. The relationship of this work to target discrimination methods for sonar applications is highlighted. [Work supported by ONR.]

1:50

5pUW4. Numerical and experimental investigations of transformations of near-field to far-field scattering measurements. John A. Fawcett, Juri Sildam, and Mark V. Trevorrow (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS, Canada, B2Y 3Z7, john.fawcett@drdc-rddc.gc.ca)

Due to the size of a target and measurement constraints, the measurements of monostatic acoustic scattering are often made in the near field of the target. For many applications, it is an estimate of the far-field target scattering strength, which is desired. In this presentation some model-based inversion techniques, which can be used to predict far-field monostatic target strengths as a function of aspect angle, are derived. In particular, a set of near-field measurements is used in an inversion scheme to compute coefficients for either a linear array of point scatterers or for a generalized Kirchhoff model of the target. The resulting model is then used to predict the far field scattering amplitudes. Simulated and experimental data are used to illustrate the accuracy of the proposed methods.

3380   J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, November 2006

Fourth Joint Meeting: ASA and ASJ   **3380**

5pUW5. Scattering in swell-contaminated seas: A semiempirical approach. Roger C. Gauss and Joseph M. Fialkowski (Naval Res. Lab., Code 7144, Washington, DC 20375-5350)

Acoustic scattering from the undersea air-sea interface is driven by its surface roughness. In modeling this scattering, 2-D spectral models such as the Pierson-Moskowitz model are used to represent the roughness contribution. However, these spectral models typically assume fully developed wind-driven seas. In practice, seas are not always fully developed and swell is often present. In general, scattering depends on the full roughness spectrum and so can be sensitive to its peak behavior. This paper describes low-to-midfrequency (< 10 kHz) bistatic scattering strength and coherent surface-loss models that are acoustically constrained by backscatter data and oceanographically constrained by coincidentally measured wave spectra. To match the latter, a semiempirical surface-wave spectral model is developed whose key parameters (strength, exponent, and peak) are expressed solely in terms of significant wave height and wind speed. It is shown that including the non-wind-wave contributions is essential to accurately modeling interface scattering in general at low frequencies, and forward scattering at all frequencies. [Work supported by ONR and PEO C4I & Space (PMW-180).]

5pUW6. Modeling reverberation time series for shallow water clutter environments. Kevin D. LePage (Naval Res. Lab., Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375), Charles W. Holland (The Penn State Univ., State College, PA 16801), and Peter Neumann (Planning Systems Inc., Reston, VA 20191)

The phenomenon of clutter in shallow water environments can be modeled from a variety of viewpoints. One approach we have taken is to model reverberation time series for heterogeneous environments with a variety of scattering mechanisms to see if some of the characteristics of resulting estimates model the characteristics of experimentally observed clutter. Here we show reverberation time series estimates for the Straits of Sicily that seem to capture some of the clutter behavior observed during the Boundary 04 experiment conducted in conjunction with the NATO Undersea Research Centre. Of particular interest is the appearance of coherent time-frequency striation patterns in both the modeled and observed reverberation, with compact time supported but broadband clutter features superimposed by the presence of environmental discontinuities. [Work supported by ONR Phase II STTR N00014-04-C-0399.]

5pUW7. Mechanism for sea bottom reverberation in long range sonar in a range-dependent continental shelf environment. Ameya Galinde and Purnima Ratilal (Northeastern Univ., 409 Dana Res. Ctr., Boston, MA 02115)

The statistically stationary background reverberation in long-range sonar data acquired in the New Jersey Continental shelf during the ONR 2003 Main Acoustic Clutter Experiment is mainly due to seafloor scattering. We compare three mechanisms or models for scattering from the sea bottom in a range-dependent waveguide with the experimental data. They are the Rayleigh-Born volume scattering model, a rough surface scattering model, and an empirical Lambertian model. Each of these models has a different decay rate as a function of range and different frequency dependence. These models are compared to the measured background reverberation as a function of range from the sonar system and at various frequency bandwidths from 300 to 1500 Hz to determine which mechanism or model for seafloor scattering best describes the data.

5pUW8. "Through-the-sensor" monostatic bottom loss and bottom backscatter measurements at mid-frequencies. Martin Barlett and Walter Brown (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713, barlett@arlut.utexas.edu)

Determinations of bottom backscatter and bottom loss parameters for the seafloor at mid-frequencies are of interest for understanding the performance of modern tactical sonar systems. A variety of techniques exist for measuring bottom scattering properties, but many of these methods require significant resources and/or time to perform. Over the last several years, a through-the-sensor approach for making bottom backscatter measurements using standard tactical sonars has been developed. Recently, these techniques have been extended to monostatic bottom loss estimates. In this presentation, a description of the overall approach to these measurements, advantages and limitations of these techniques, and examples of through-the-sensor bottom backscatter and bottom loss measurements made from data collected through a tactical sonar system will be given.

3:05–3:20   Break

5pUW9. Validating physics-based clutter from seafloor roughness and biologic scattering. Juan Arvelo, Charles Monjo, Bruce Newhall (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099), Adam Frankel (Marine Acoust. Inc., Annapolis, MD 21401), Richard Love (Fort Worth, TX 76179), and William T. Ellison (Marine Acoust. Inc., Litchfield, CT 06759)

Two physics-based clutter models were developed for comparison and to study the merits and limitations of both approaches. The stochastic approach [Newhall and Arvelo, J. Acoust. Soc. Am. **118**, 2041 (2005)] makes use of perturbation theory, to relate the mean seafloor scattering strength to the mean surface wavenumber spectrum, and the extrapolation of this spectrum to the Bragg wavenumber to generate the clutter distribution from the seafloor roughness. The semi-deterministic approach [Monjo and Arvelo, J. Acoust. Soc. Am. **118**, 2041 (2005)] is based on 2-D power-law fractal realizations of the ocean floor, coherent modal addition, and the microscale bathymetric slopes. Biologic scattering was later included where spatial distribution and nonstationarity of schools of fish and individual marine mammals are realistically represented [Frankel *et al.* J. Acoust. Soc. Am. **119**, 3437 (2006)] through individual-based modeling of their movements, school size, and target strengths. Clutter distributions from both approaches are compared against each other, for selected test cases, and against clutter measurements. [This effort was funded by the Maritime Sensing Program of the Office of Naval Research (ONR Code MS 321).]

5pUW10. Ocean bottom scattering: characterization with chirp sonar. Jorge Quijano, Lisa Zurk (Portland State Univ., P.O. Box 751-ECE, Portland, OR 97207, zurkl@cecs.pdx.edu), Altan Turgut (Naval Res. Lab/DC, Washington, DC 20375), and D. J. Tang (Appl. Phys. Lab., Seattle, WA 98105)

Shallow water acoustic propagation in the mid-frequency range (~ 1–10 kHz) is strongly affected by scattering from the ocean bottom. In general, this scattering is a combination of rough surface scattering from the water-sediment interface and sediment layers, as well as volume scattering from embedded inhomogeneities. It is necessary to develop an understanding of the spatial and temporal characteristics of the acoustic field to optimize Navy sonar performance. Although there has been a growth of research activity in this area, the validity of different modeling approaches still remains an open research question. One promising tool for bottom sediment characterization is a chirp sonar. These sonars can produce high-quality data in which time of arrival is often associated with location of subbottom sediment layers to produce sediment profiles. However, features of the acoustic scattering such as multiple scattering or particle resonances can have complicated signatures that are not straightforward to

interpret. In this work, the effect of these signatures on the chirp sonar data is studied through application of a numerical FDTD model applied to a variety of ocean bottom geometries. Results obtained from the chirp sonar system are presented and compared with ocean bottom scattering models.

3:50

5pUW11. Using the ultrasound standing wave in vibroacoustography for surface roughness imaging applications. Farid G. Mitri and Mostafa Fatemi (Dept. of Physiol., and Biomed. Eng., Ultrasound Res. Lab., Mayo Clinic College of Medicine, 200 First St. SW, Rochester MN 55905)

The surface roughness (of about 20 microns) of an acrylic block immersed in water was measured using the acoustic emission pressure field (in the kHz range) induced by the dynamic radiation force of ultrasound (2 MHz). The dynamic radiation force is produced by mixing two beams of different frequencies to excite the probed object. The acoustic emission is a result of object vibration at the difference frequency of the incident ultrasound beams and is detected by a low-frequency hydrophone placed nearby the block. Due to the presence of the block, an ultrasound standing wave field is generated as a result of multiple reflections between the transducer and the block's surface. The acoustic emission by the block's surface varies according to its relative position to the transducer. This variation is used to form an image of the surface roughness at two difference frequencies $\Delta_f = 9.9$ and 22.4 kHz. To have a one-to-one map between the surface roughness and the image, the surface variation has to lie within the distance from a maximum to its nearest minimum in the standing wave. Images obtained here demonstrate that vibroacoustography may be used as a powerful tool for the nondestructive inspection and imaging of surface roughness [Mitri *et al.*, Appl. Phys. Lett. **88**, 234105 (2006)]. [Work supported by NIH.]

4:05

5pUW12. Tower-based breaking wave noise measurements. Steven L. Means and Jeffrey A. Schindall (Naval Res. Lab., Code 7120, 4555 Overlook Ave. SW, Washington, DC 20375, means@wave.nrl.navy.mil)

A tower-based sea surface noise experiment began collecting data January 2006 in the shallow waters (25 m) approximately 75 km off the coast of Savannah, Georgia. A 32-phone, three nested-aperture, vertical hydrophone array was deployed nominally 100 m from a Navy tower that stands 50 m above the water surface. A high-resolution video camera was mounted near the top of the tower along with a dual-polarized marine-band radar to record the location, size, and lifetime of the surface expressions of the breaking waves above of the array. The array is cabled back to the tower for power and signal collection. The tower is microwave-linked to shore for internet-based control and data retrieval. To date, measurements have been made in wind speeds ranging from 5–21 m/s and wave heights of 1–3.4 m over a 6-month period. Empirical relationships between the time-frequency structures of the generated noise, obtained on the endfire beam of the array, and the size of the surface expressions of the breaking waves will be given for a range of environmental conditions. [Work supported by ONR base funding at NRL.]

4:20

5pUW13. Numerical simulation of noise generated by wave breaking in a tank. Xuemei Chen, Steven L. Means, Bill G. Szymczak (Acoust. Div., Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375, xchen@wave.nrl.navy.mil), and Joel C. W. Rogers (Brooklyn Polytechnic Univ., Brooklyn, NY)

A coupled hydro-acoustic source model in relating the physical parameters of wave breaking to the source quantities has been developed. The physical processes of wave formation and breaking are modeled using a generalized hydrodynamics formulation with the initial wave profile cal-

culated by boundary integral method. Hydrodynamic parameters, such as pressure variations and air cavity shapes, etc., will be provided through the simulation. In the acoustic simulation, an algorithm has been developed in handling wave propagation in irregular regions, such as the bubbly liquid generated by wave breaking. To validate the modeling, an experiment on wave formation, propagation, and breaking is carried out in a wave tank. The waves are generated mechanically with a computer controlled vertically oscillating wedge. Favorable agreement is found upon comparing prediction and measurement. [Work supported by ONR.]

4:35

5pUW14. Long-term trends in ocean noise. D. Benjamin Reeder, Rommel Pucan, and Curtis A. Collins (Naval Postgraduate School, 833 Dyer Rd., SP-222 Bldg. 232, Monterey, CA 93943)

The effect of ocean noise, particularly from anthropogenic sources such as Navy sonar systems, on marine mammal hearing, behavior, and reproduction is of great concern. However, relatively few studies exist to make a definitive conclusion regarding rising ambient noise levels, due largely to the dearth of information over sufficiently long time periods for a sufficient number of geographical locations. A few studies have estimated that ocean noise is rising at a rate of approximately 3–5 dB per decade at some frequencies in some locations. This study is based upon ambient noise measurements from U.S. Navy sonobuoys deployed by operational units between 1993 and 2004. Data and analysis are presented for three areas of the ocean—the Northwestern Atlantic Ocean, the Greenland-Iceland-United Kingdom (GIUK) gap, and the Mediterranean Sea. Analysis of this data demonstrates a less significant increase, and in some cases, a decrease, in ocean noise for these three ocean areas at frequencies between 50 and 2000 Hz. Ocean noise levels in this study are comparable to measured ocean noise levels reported in the literature during the 1960's.

4:50

5pUW15. Vessel noise measurements underwater in the Haro Strait, WA. Scott Veirs (Beam Reach Marine Sci. and Sustainability School, 6537 16th St. NE, Seattle, WA 98115, scott@beamreach.org) and Val Veirs (Colorado College, Colorado Springs, CO 80903)

In the recent listing of southern resident orcas as endangered under the Endangered Species Act, disturbance by vessels is cited as a risk factor for the population and a factor in the designation of critical habitat. To help determine whether vessel noise underwater is disturbing orcas as they use sound to navigate, forage, or communicate, the source level of common large vessels transiting Haro Strait was measured. Since large vessels are the dominant source of underwater sound for the southern residents (based on water volume ensonified and duration of exposure at a given receive level), it is convenient that all commercial vessels longer than 65 ft are now required to broadcast their location, speed, and other data in real-time. The automatic identification system (AIS) broadcasts data every 2 s via VHF that can be received and decoded with specialized radio receivers that are now coming on the market. AIS was used to log range, speed, and relative bearing while simultaneous recordings were made of the received sound pressure levels using calibrated hydrophones and photographs were taken. From these observations, the broadband and spectrum source level of each vessel was calculated as a function of bearing relative to vessel axis.